

Essentials for science in a devolved Scotland

A newly elected parliament and executive provides opportunities for Scotland not only to imitate the best of Westminster, but also to do better in its handling of Scottish research and public controversies.

Scotland is no stranger to self-government. It has had a parliament since 1235, long before the union of Scottish and English parliaments in 1707. The legacy of that separate identity remains, for example, in Scotland's law and its education system. This tradition of independent thinking should continue following last week's elections to a new parliament in Edinburgh, Scotland, and a national assembly in Cardiff, Wales. The likelihood of coalition governments in both legislatures and a strong representation from political parties that want independence from the United Kingdom suggest that there could be exciting times ahead.

An area where fresh ideas are sorely needed is in the relationship between science and government, following the collapse in public confidence in official channels of science advice during the debacles over BSE and genetically modified food. Scientists need to encourage parliamentarians to develop a deeper awareness of the strengths and limitations of science, and to move away from the view that science is a fixed, unchanging body of knowledge, which can always be relied upon to underpin complex policy.

One of the biggest tests for both the Scottish parliament and the Welsh assembly will be in the degree of pre-legislative scrutiny granted to the public and to special-interest groups.

They will also need to be responsive to the perception by the public of threats to the environment and health. To be fair, the UK government is well aware of these issues and will soon announce the results of its survey of public perceptions of the regulation of bio-science and biotechnology. These results are expected to feed into a parallel review of the government's many advisory committees.

While they seek to be creative, the new legislatures should not overlook some of the more successful aspects of the UK parliament at Westminster. One example is the science and technology select

committee, which scrutinizes the work of the executive. In the context of many competing pressures from powerful interest groups, no modern assembly can afford to be without such a committee.

Many of the issues facing the new administrations — for example, the regulation of fish stocks or concentrations of airborne particulate matter — will have a strong scientific dimension. The Westminster parliament relies on its Parliamentary Office of Science and Technology for independent advice on these issues. The Scottish parliament, at least, will need an analogous mechanism.

Scotland has a significant number of strengths in science, and its researchers are at the forefront of UK government efforts to commercialize research (see page 97). The Royal Society of Edinburgh and the UK Scottish Office are right to recommend against a division of the UK research-council budget, as this would have left Scottish university researchers competing for fewer funds. Scotland's agricultural and biological research institutes, by contrast, are understandably less enthusiastic about the new funding arrangements. They will now have to lobby parliament for their budgets, when in the past they were funded directly by the Scottish Office. Some of their concerns, however, could be alleviated if the parliament chose to set up a smaller executive office of science and technology to distribute its own funds and to protect them from direct political control.

In 1457, the Scottish parliament decreed that golf and football were not appropriate forms of recreation for men under 60, and banned them in favour of archery, considered a more effective way of ensuring personal fitness and self-defence. The new parliament need not be so radical. But there is no reason why it cannot be innovative, not least in its dealings with science. □

Impure, maybe, but exemplary

Whatever its motivations, a drug company's big gesture towards AIDS should be imitated.

The news that a leading US pharmaceutical company, Bristol-Myers Squibb (BMS), is planning to spend \$100 million on a programme of AIDS research, treatment and education targeting women and children in southern Africa (see page 96) is to be welcomed.

It is true that an \$18 billion company like BMS can readily afford such an investment. Cynically, one might suggest that the move has been timed to coincide with preparations by the World Health Assembly to take a vote that would encourage governments of poorer countries to ignore patents on expensive AIDS medicines and license the in-country manufacture of cheaper generics. It can be seen in a longer-term perspective: by adopting the moral high ground ahead of its competitors, BMS may be positioning itself well for the vast demand for anti-AIDS drugs that is unfortunately likely to open up in the continent in the years ahead.

But the fact that BMS's largesse may have been informed by bottom-line considerations doesn't remove the value of its gesture. The money will be used to train doctors, help AIDS orphans and fund \$44 million of clinical research. Its arrival calls attention to the fact that four out of five AIDS deaths worldwide have occurred in sub-Saharan Africa, which is home to about two-thirds of the estimated 34 million people currently afflicted by AIDS.

Against that backdrop, one can hope that the BMS initiative will encourage other drug companies to play a part in addressing Africa's AIDS crisis. Future initiatives should involve all the major pharmaceutical companies, as well as national health authorities, the funding bodies responsible for health care, and perhaps even non-health organizations that stand to lose financially from an unchecked AIDS epidemic, with the goal of making essential medicines available to Africans at an affordable cost. □



Quake-spotting 'telescope' may map North American continent

[SEATTLE] A consortium of 90 US universities is planning an unprecedented new seismology facility to map North America. It will use a network of portable devices that its supporters describe as a form of Hubble telescope looking deep into the Earth.

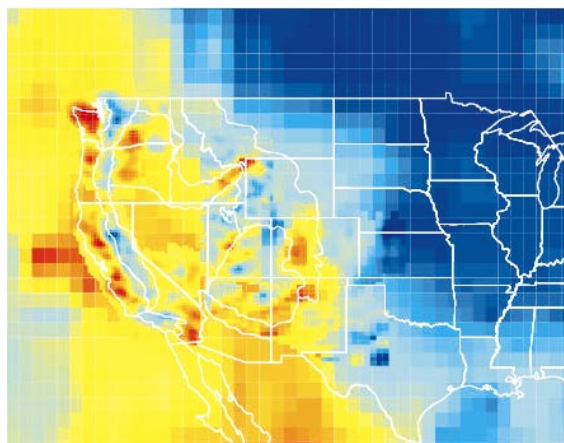
Funding for the experimental facility, currently called USArray, is being sought by Incorporated Research Institutions for Seismology (IRIS). The consortium oversees major capital expenditure for member universities' research projects from agencies such as the National Science Foundation (NSF).

"The heart of the facility is a transportable, dense array of seismometers, which will improve the quality of seismic images of the continental lithosphere and deeper mantle by an order of magnitude," says a report from a workshop held by IRIS in March in Albuquerque, New Mexico.

The report says that a series of deployments of such an array would cover the contiguous United States in ten years, and result in a uniform data set that would be ideal for imaging the continent at high spatial resolution. The array will use new communication technologies, it adds, allowing data to be made available in real-time on the Internet.

The NSF has already included the anticipated costs — estimated to be tens of millions of dollars — in its long-range budget plans. IRIS intends to send a report on USArray to the foundation in the next month. NSF officials say they expect to brief the advisory National Science Board on the project in the autumn, and seek funding from Congress during hearings for the 2001 budget.

To muster support, IRIS scientists briefed representatives from the Council of the National Seismic System, a group of some 30 seismology networks that held its annual meeting in Seattle last week.



Looking into the depths: this composite image of upper mantle P-wave seismic structure, some 100 km below the western United States, shows more variation than seismologists had previously suspected. Red areas show the partially molten, low-velocity mantle; blue represents subsolidus, high-velocity mantle.

Seismologists say the array could provide researchers with a rich catalogue of information about the structure and formation of the continent, while offering insights into such topics as how surface water flows, the safety of sites for waste disposal, and more explanations of mineral and oil deposits.

"It is an opportunity to really understand the continent we live on," says IRIS's president David Simpson. He calls the experiment "a major challenge" — a reference to the logistics of collecting and managing the data.

If USArray is funded and set up, proponents hope to coordinate their efforts with seismologists in Canada and Mexico to expand mapping of the Earth's crust.

USArray would be the biggest project undertaken by IRIS (which has an annual operating budget of about \$11 million) since it was created 16 years ago to combine member institutions' resources to advance seismology research cost-efficiently.

Seismologists touted the facility's opportunities enthusiastically at the Seattle meeting. But they are realistic about the federal government's funding process. USArray

faces "a long road ahead" before it is funded, they noted, and a delay in creating the facility is a distinct possibility. It also faces obstacles from within and without seismology ranks.

Steve Malone, for example, a University of Washington seismologist who chaired the Council of the National Seismic System for the past year, told the assembly of more than 50 scientists that it would require seismologists to think more collectively than in the past. Malone warned of the importance of avoiding turf wars, data-hogging and seismologists' tendency to consider an earthquake that they identify as "my earthquake".

Anne S. Meltzer, a seismologist at Lehigh University, Pennsylvania, who chairs IRIS's executive committee, added that USArray could "expand the culture of shared and coordinated resources within Earth science".

The USArray will require an unprecedented number of seismometers from the few companies that make the very broadband, high dynamic range seismometers that are required. At least 400 seismometers will be required for the portable array — passive equipment that sits on temporary pads, recording waves from near and distant earthquakes to image the Earth's crust and mantle.

According to a report on a USArray workshop held in March by Peter Shearer, of the Scripps Institution of Oceanography in San Diego, and colleagues, the facility "will provide an opportunity for local universities and research groups to target specific geologic problems". A regional pre-deployment workshop would be held for each stage, to help plan experiments and coordinate educational outreach efforts.

"Whenever a new source of seismic data becomes available, unexpected discoveries have often overshadowed the original purpose of the data collection. Almost certainly USArray will yield many such bonuses," says Shearer and his colleagues.

Rex Dalton

Search for a key to secrets of the Universe

[LONDON] Scientists at the Stanford Linear Accelerator Center in California started a search on Tuesday (11 May) for evidence of short-lived subatomic particles known as B-mesons, from collisions of electrons and positrons.

The B-mesons and their anti-particles will be detected using a 1,200-tonne particle detector known as BaBar, positioned where the electron and positron beams meet.

The experiment is designed to find clues as to why the Universe has more matter than antimatter.

Scientists believe that microseconds after the Big Bang, equal amounts of matter and anti-matter should have annihilated each other. "Mysteriously, a bit of matter was left over," says Ken Peach of the UK Rutherford Appleton Laboratory, one of the centres involved in the

experiment. "By studying particles and their anti-particles we hope to see the tiny differences in behaviour between matter and anti-matter, which will help explain the huge imbalance in the Universe," adds Peach.

BaBar, which cost \$300 million, is the result of collaborative work by more than 500 physicists and engineers from 10 countries.

Ehsan Masood

'Sell the message of research', Europe's cell biologists urged

[BOLOGNA] European scientists should talk more about the importance of their work — to both the public and politicians — if they are to be as successful in attracting funding as their colleagues in the United States.

This was the message transmitted by some of Europe's top cell biologists at a meeting of the European Cell Biology Organization (ECBO) in Bologna earlier this week.

According to Jacopo Meldolesi, head of neurobiology at the DIBIT research institute in Milan, European scientists are traditionally reluctant to band together and lobby politicians — particularly the European parliament — about the importance of basic research.

Meldosi says this is partly because the scientific communities are fragmented along national lines. Many scientists believe that a lack of lobbying force has eroded the role of basic research in the European Union's Framework programmes of research.

Last month, for example, at a meeting organized by Euroscience, the 'grass roots' association of European scientists, Members of the European Parliament criticized scientists in general for a tendency towards 'renationalization', and for not getting together and lobbying alongside industry and pressure groups such as Greenpeace (see *Nature* 398, 646; 1999).

Kai Simons, senior scientist at the European Molecular Biology Laboratory in Heidelberg, says that political lobbying is one of the aims of the newly formed European Life Sciences Organization (ELSO). This body is intended to take over from ECBO, a federation of national societies, which has failed to attract large numbers to its annual meetings (see *Nature* 393, 615; 1998).

Despite its strong scientific programme, fewer than 900 scientists attended the meeting in Bologna. This is in contrast to meetings of the American Society for Cell Biology, which regularly attract 10,000 participants.

Simons hopes for up to 4,000 participants at ELSO's first meeting in Geneva next year. The organization is being modelled on the American society, he said. Its meetings will be cheap to attend, and it will also organize political lobbying, "but only if we get the mandate from members to do so".

Peter Rigby, director of the Chester Beatty Cancer Research Institute in London, says that scientists who are funded by charities should talk not only to politicians but also to the street collectors for those charities. The latter "need to know how the money is spent — even if the concepts are hard to put across — in order to keep up their motivation", Rigby says. **Alison Abbott**

US stem-cell pioneers buy 'Dolly' cloning company

[WASHINGTON] The US biotechnology company that financed the isolation of human embryonic stem cells has joined forces with the cloners of Dolly the sheep in a bid to combine the two technologies and speed the development of therapies for degenerative diseases.

Geron Corporation of Menlo Park, California, which announced the stem-cell work last year, said last week that it has bought Roslin Bio-Med, a company that was set up by the Roslin Institute, near Edinburgh, to develop its cloning technology (see *Nature* 396, 104; 1998).

The price was 2.1 million shares of Geron stock — worth about \$26 million at the time of the agreement. Geron will spend an additional \$20 million in the next six years funding research at Geron Bio-Med, as the renamed company is called.

In addition to the purchase price, the Scottish institute will continue to benefit from royalties from new intellectual property arising from the research funded by Geron, \$4 million of which will go to Roslin's animal genome research.

The research to be conducted by Geron Bio-Med aims to unravel how the cytoplasm of an egg 'reprogrammes' an adult cell nucleus, and to confer the same ability on a store of generic, enucleated cells that are not eggs. The goal is to circumvent the need for human oocytes as recipient cells in the cloning process, so avoiding both the ethical questions surrounding the cloning of human embryos, and the practical obstacle of a shortage of human oocytes.

The ultimate goal, according to Thomas Okarma, Geron's vice-president of research and development, is to inject a patient's own DNA into such a 'generic reprogramming cytoplasm', then to direct the cell created in this way to form the tissue that is needed. In Parkinson's disease, for example, these would be dopamine-generating neurons. Because a patient is his or her own DNA donor, rejection should be avoided.

Geron has already cloned critical parts of human telomerase, an enzyme that allows cells to keep replicating beyond the time they would naturally have stopped — an important capability if transplanted cells are to replace those that have died or are not functioning properly.

"There is incredible synergy under one roof now with these three technologies, and with the collaboration with the Roslin Institute, which is obviously the premier animal genomics centre of the world," says Okarma.

"We feel that [the Geron-Roslin partnership] will begin to open doors towards



Wilmut: seeking to reprogramme cells.

cell-based therapies for a whole variety of degenerative diseases," says Ian Wilmut, the Roslin Institute scientist who cloned Dolly, and who will be chief scientific officer for Geron Bio-Med.

"The long-term aim is to re-program human cells without using eggs or creating embryos," says Wilmut. He will run the research effort with John Clark, head of the molecular biology

division at the Roslin Institute.

Some remain sceptical about the feasibility of producing healthy cells — or animals — by cloning, and so doubt that the Geron-Roslin collaboration will succeed. "Taking an adult cell and reprogramming it is infinitely difficult," says Norton Zinder, a biologist at Rockefeller University in New York. "I don't think it will work in the laboratory."

Zinder points out that *The Lancet* published a paper on 30 April by French scientists reporting the death of a cloned calf at seven weeks from anaemia. Its lymph nodes and spleen were not properly developed — an indication, he says, of the type of unexpected problem that could arise.

Investors apparently share some wariness about the new venture. Geron's stock closed at \$11.75 last Friday (7 May), down from \$12.37 on 3 May, the day the Roslin deal was announced. This was in contrast to last November, when the company's stock soared briefly to \$24.50 on the news of the stem-cell isolation.

But others remain upbeat. "This deal will ensure that British scientists will play a key role in developing therapies that are potentially amongst the most exciting in human medicine," says Grahame Bullfield, director of the Roslin Institute.

Wilmut holds 83,000 shares in Geron Bio-Med, formerly Roslin Bio-Med. He says he was required by 3i Group, the original investor in Roslin Bio-Med, to purchase the shares as a condition of the launch of the company in 1998, to demonstrate his long-term commitment to its projects. The shares — not accessible to him for a year — were worth \$975,000 last Friday. "Who knows what they will be worth a year from now?" says Wilmut. **Meredith Wadman**

Battery fault ends X-ray satellite mission

[MUNICH] Astronomers around the world were last week lamenting the likely loss of Abrixas, a small German-built X-ray satellite whose batteries failed two days after a successful Russian launch.

Some were also suggesting that the failure has raised doubts about the apparently high level of risk associated with so-called 'smaller, faster, cheaper' science missions. These are favoured by space agencies in Europe and the United States as a way of introducing more flexibility into their launch schedules (see *Nature* 389, 899; 1997).

Abrixas was to have carried out the first complete broad-band all-sky survey with imaging telescopes in the medium-energy X-ray range, extending by an order of magnitude the survey performed by its predecessor ROSAT, which was shut down last year.

Abrixas was seen as a pathfinder for future large international X-ray missions, including the European Space Agency's XMM, the US space agency NASA's AXAF — recently renamed the Chandra X-ray Observatory — and Japan's Astro-E.

Scientists on these missions, each scheduled to be launched within the next nine months, had been hoping to study interesting X-ray sources identified by Abrixas.

But ground contact with the satellite was lost two days after its launch. At least one of its eleven battery cells appears to have been burnt out by inappropriate power input from batteries used to support the launch. This means that, although the scientific instruments seem to be working, the information they gather cannot be transmitted to Earth.

Engineers will be able to analyse the full extent of the damage during a six-day period



Lost in space: but calibration data from the X-ray telescopes on Abrixas may still be useful.

at the end of next month when Abrixas moves into full sunlight and its solar panels become operational. But it seems unlikely that the battery system can be restored or the satellite's scientific mission fulfilled.

"This is not only a major disappointment to our German colleagues, but also a setback to the international astronomy community," says X-ray astronomer Ken Pounds, professor of space physics at Britain's Leicester University. "We were looking forward to Abrixas pointing the way to the best use of the upcoming international missions."

Claude Canizares, director of the Center for Space Research at the Massachusetts Institute of Technology, and principal investigator of a Chandra instrument, agrees that the situation is "very sad". Not only would Abrixas have helped locate the best targets for Chandra to focus on, he says, but "it is also nice to have relatively low-cost, small, fast and clever missions in one's portfolio".

Canizares says it is disconcerting that "so many small missions seem to be failing". Only a few weeks ago, WIRE, a small NASA satellite that would have surveyed the infrared sky, failed because its protective cover was lost after launch (see *Nature* 398, 100; 1999). Solar radiation heated the spacecraft and its store of cryogenic fuel was boiled off within a few days. This ended the mission, as infrared detectors can only function at very low temperatures in space.

Canizares suggests that the Abrixas failure is likely to fuel debates on how to reconcile speed with quality control.

Martin Turner, professor of astrophysics at Leicester University, and principal investigator on an XMM instrument, echoes the concern about faster, cheaper missions. "The Abrixas people did everything right: they completed to launch within three years, at low cost [DM40 million]," he says. "Yet they were let down by a low-tech item, the batteries."

Abrixas, like most scientific satellites, was not insured. Although the error is believed to be a design fault in the battery system, which was provided by the Bremen-based company OHB, a spokesman for the German space agency DLR, says that, if tests in June prove that the mission is dead, "it is not clear if any form of compensation could be expected".

Joachim Trümper, director of the Max Planck Institute for Extraterrestrial Physics in Garching and head of the Abrixas project, is trying to remain positive. Although Abrixas may never yield scientific data, he says, calibration work on its novel X-ray camera may be done during the satellite's brief sunlight period. The high-tech camera will also fly on XMM.

Alison Abbott

French unions hopeful of delaying reforms until after consultation

[PARIS] Trade unions representing French scientists were increasingly confident of victory this week in their demand that the government should not undertake any new reforms before the completion of a national consultation on research in July. They appear to have won the support of the prime minister's office.

The unions had called on researchers to support a national day of protest on 11 May to secure this commitment. But on Monday it seemed that the science ministry had bowed to their main demand, postponing an interministerial meeting on research originally scheduled for 18 May.

The unions had argued that the ministry's decision to proceed with the meeting suggested that it would ignore the consultation and push ahead with its own reform plans (see *Nature* 399, 4; 1999). The ministry countered that the meeting would

deal only with strategic research themes and industrial research.

A science ministry spokesperson says the meeting will only be postponed by a matter of days, but Herbert Maisl, science adviser to Lionel Jospin, the prime minister, says it is "cancelled until further notice".

The prime minister's office is now setting the timetable for reform. Jospin himself commissioned the parliamentary mission, following Socialist Party concern that science minister Claude Allègre's attempts to impose reform were opposed by the research and higher education community.

The unions' claims that the science ministry might only pay lip service to the consultation seemed to be strengthened last week by a leaked preparatory document for the meeting. The document, released by the unions on Friday, largely covers the remit of the parliamentary mission, which is led by

two Socialist members of the national assembly, Pierre Cohen (Haute-Garonne) and Jean-Yves Le Déaut (Meurthe-et-Moselle), a former head of the Parliamentary Office of Scientific and Technological Choices.

Le Déaut plays down the document's significance, arguing that it dates from several months ago. This is confirmed by Jean-Yves Mérimond, vice-chancellor of the University of Strasbourg and chairman of the research committee of the Conference of University Vice-Chancellors.

Le Déaut says Allègre gave him his personal guarantee last week that "no decision would be taken within the remit of our mission until this is completed".

Cohen says any attempt by the interministerial meeting to answer the questions of the mission will not be tolerated. The decision of the prime minister's office suggests that it shares this view.

Eric Glover

Funding penalty for cross-boundary work

[LONDON] University departments whose 'structures' do not match those followed by Britain's Research Assessment Exercise (RAE), the regular initiative used as a basis for calculating funding allocations to universities, appear to get systematically lower research ratings.

According to an independent report on interdisciplinary research, nearly a quarter of university departments in the last Research Assessment Exercise either split their researchers between assessment panels, or sought cross-referral between panels.

The report found that such 'boundary critical' submissions received on average a 0.5-point lower rating than non-critical submissions. The RAE rates departments on a 5-point scale, and there are serious financial consequences for those that receive a reduced rating, particularly at the top of the scale.

The report, by consultancy Evaluation Associates, was commissioned by the four regional higher education funding bodies in preparation for the next Research Assessment Exercise, due to take place in 2001. It says of submissions that were cross-referred and split between panels that "our concern is

that assessment in these circumstances is inherently difficult for panels. They are not considering entire departments but potentially arbitrary subsets of departments".

But the report also says that it found no evidence that RAE inhibits interdisciplinary research, despite a widely held conviction by researchers that it does. It did find a lack of consistency in the treatment of interdisciplinary research by individual panels, however.

"I've never had any problem with [the RAE] and I don't think it is a problem," says one interdisciplinary researcher involved with the next RAE. "In the area I operate, there is no disadvantage and I am extremely involved in interdisciplinary research."

Bahram Bekhradnia, director of policy at the Higher Education Funding Council for England (HEFCE) — which was responsible for the last exercise — says that interdisciplinary work "has always been one of the more difficult aspects of the exercise which is a discipline-based process. This was one of the reasons for the review."

He said that the report gave general comfort to interdisciplinary researchers, and that the funding council "expected to imple-

ment" the specific recommendations. The report recommends new mechanisms for boundary critical submissions, and monitoring mechanisms to "ensure the effectiveness of cross-referral".

Anita Jackson, a member of the RAE 2001 team, says that the assessment process is "evolutionary", and that the next exercise will take on board the recommendations of the report. "We will ask departments to say if their units do not fit. We have put in mechanisms where anything cross-referred will happen straight away — which was not what happened in the past."

It remains to be seen whether this will impress departments that feel they were badly rated in the last exercise. "We are aware that if anyone feels aggrieved they can have a judicial review," says David Pilsbury of HEFCE. But he points out that "it happened once before, and the outcome was in our favour".

The report finds that the RAE has encouraged researchers to concentrate on the quality of their research, but that they are doing more research that produces results in the short rather than long term. **Natasha Loder**

Australia boosts medical research, but keeps other budgets level

[CANBERRA] For the first time in his four budgets, the Treasurer of Australia's centre-right Coalition government, Peter Costello, has highlighted the importance of science and research in his speech to Parliament, delivered in Canberra on Tuesday (11 May).

As foreshadowed by Prime Minister John Howard last week, medical research was given star billing in the budget, and is scheduled for a boost of A\$614 million (US\$404 million) over six years.

But other research budgets have been kept roughly at this year's levels. And the Science and Technology Awareness Program, which, among other activities, supported National Science Week, and cost A\$3 million in the past year, is being shelved "pending review".

As a result of the extra money for medical research, the National Health and Medical Research Council will have doubled its current budget of A\$161 million by the year 2005. "We as a government are very excited [about this]," said Costello. "It will put us at the cutting edge to really develop new industries."

The boost to medical research is the first outcome of a year-long review of health research that was chaired by businessman Peter Wills. Further recommendations from the panel are expected to be approved by the government shortly.

But although the medical and other



Costello: 'puts us at the cutting edge'.

measures may halt the declining support for research that has occurred throughout the Coalition's first three-year term, overall Australia is not matching the scale of increases for science seen in several other countries, such as Canada, with similar size economies.

One disappointment was that a special programme for biotechnology, which had been tipped to receive an extra A\$60 million, will only receive A\$17.5 million over two years — and this is earmarked not for research but for "development of a comprehensive new biotechnology strategy".

But the government promised to use some of the funds to set up a senior ministerial council "to manage the biotechnology agenda" and establish a statutory office to regulate the industry.

University infrastructure, long a subject of complaint from academic circles, is scheduled to receive an increase of A\$93 million, to bring total funding to A\$288 million over three years. But this merely reverses cuts foreshadowed previously, and seems to be balanced by falls in other education programmes.

Funding of research grants remains steady pending the conclusion of a major review of the Australian Research Council.

Nick Minchin, the Minister for Industry, Science and Resources, has negotiated level funding for three national agencies operating under his aegis, the Commonwealth Scientific and Industrial Research Organization, the Australian Nuclear Science and Technology Organization (to which he gave final approval for a new research reactor last week), and the Australian Institute of Marine Science.

The annual Australia Prize, costing A\$800,000, will be continued to 2000–01 only because of its "significant lead times", according to a budget document.

Bob McMullan, shadow minister for industry and technology in the Labor Opposition, attacked the government for "hiding overall decreases" across several portfolios with stakes in science and technology by not releasing the annual science and technology budget statement on Tuesday.

Brian Anderson, president of the Australian Academy of Science, says the academy is "delighted with the result in medical research". But he says universities are still underfunded, and "there are no measures to reverse the downward trends in research by industry". **Peter Pockley**

NSF told to ease up cost-share demands

[WASHINGTON] Programme officers at the US National Science Foundation (NSF) have been told to stop using their considerable clout with principal investigators to negotiate down the size of research grants. This tactic shifts part of the cost of the research from the NSF to the university receiving the grant.

Last week, the National Science Board, the NSF's governing body, approved a new cost-sharing policy that prohibits the reduction of a grant without a corresponding reduction in the scope of the proposed work. It also requires cost sharing to be declared up-front, and prohibits its use as an unwritten criterion for selecting grants.

"There is a lively interest among members of the community and the board with regard to cost sharing," says John Armstrong of IBM, a member of the board. "We'll be taking a substantial step forward with this policy."

University administrators and researchers have been growing alarmed about the practice — which they regard as 'cost shifting' from the government to the universities — because of its cost and the tension it creates between grant recipients and administrators.

Administrators blame investigators for surrendering money to the government in unauthorized negotiations with the officials who administer the grants. Researchers may increase their chances of winning a grant if they can reduce their claim, because the programme managers have limited sums of money to support particular fields.

Officials admit that the issue has been around for years. But they say it has only emerged as a major bone of contention as government auditors have begun to unearth some of the dark secrets of university finance, prompted by concern in the Congress.

Every few years, auditors descend on every research university to check the amount spent directly on research, and that spent indirectly to support research. The ratio between the two — the overhead rate — critically determines how much money the government should pay to the institution to cover the costs of commissioned research.

The auditors insist that the time spent by investigators on research projects should be treated as a direct cost, even if the investigators have not claimed the full amount from the funding agency that supports them. If, for example, a principal investigator agrees to take no money from the NSF to cover time spent on a project, but subsequently spends half of his or her time on it, the auditors will insist that the time was spent on research.

If such a practice is repeated across dozens of grants, it increases the university's direct research costs, and therefore reduces the overhead rate it can claim. One institution in Massachusetts recently repaid \$12 million to the government after an audit, and

another in New York paid back \$6 million.

Two weeks ago, President Bill Clinton promised to make arrangements for cost sharing more transparent across the government, as part of a new partnership with the universities (see *Nature* 399, 3; 1999). The agency most widely accused of shifting costs to the universities is the NSF, the largest supporter of non-biomedical research at US universities. (Grants from the National Institutes of Health usually cover researchers' salaries, so disputes about cost sharing are less likely.)

"The NSF stands out because its programme managers are known to push hard to get universities to pay for a larger share of research costs," says Milton Goldberg of the Council on Government Relations, which represents universities in Washington.

The issue has dominated recent regional meetings at which researchers and university administrators met NSF officials to discuss grievances. "It is like a powder keg going off" when cost sharing is raised, says one official.

The new policy is intended to quell this discontent and address concerns elsewhere in the government that NSF cost-sharing practice is driving down overhead rates to levels that do not reflect the full cost of research. A White House official points out that artificially low overhead rates will save money in the short term, but will damage research in the long term by deterring universities from investing in equipment and buildings.

The policy states that any cost-sharing requirement in an NSF programme will be "clearly stated in the programme announcement", and that negotiations on cost sharing will be decoupled from the merit review process that determines who receives grants.

Standard NSF grants should require no cost sharing beyond an arcane statutory requirement for institutions to bear one per cent of project costs, says the policy statement. It adds that programme managers "may not negotiate or impose cost sharing or



other institutional commitments" during the final discussions they hold with principal investigators to determine grant amounts.

In these final negotiations, the statement concludes, "any reduction of ten per cent or more from the amount proposed should be accompanied by a corresponding reduction in the scope of the project, unless the programme officer, principal investigator and institution clearly agree" that the cut can be made without the institution incurring extra costs.

But Mary Clutter, NSF associate director for life sciences, one of the officials whose staff stand accused of squeezing investigators during grant negotiations, doubts that the policy will make much difference to the administration of ordinary NSF grants. "They're not supposed to be doing these negotiations even now, so I don't know if [the policy] will have that much impact," she says. But she says she likes the policy because it will make it clear what cost sharing is expected.

Response in Washington was sceptical. "It is likely to make a small difference," says Goldberg. "Success is problematic because program officers are pretty good, and there are ways of getting round it." **Colin Macilwain**

Future of Spectrum X mission still cloudy

[MUNICH] Representatives of national space agencies involved in the Spectrum X collaboration, an ambitious and long-delayed X-ray astronomy mission, will meet in Moscow next week to discuss prospects for the project, now due for launch in 2001.

Space agencies have already built instruments for the mission, which was conceived by the former Soviet Academy of Sciences, for a mid-1990s launch.

But the building of the main spacecraft, by Russia's satellite manufacturer Lavochkin Association, has never been completed. The problem is believed to be financial.

"Space agencies have been frustrated by

unclear pictures of what is happening," says Paul Merdin of the British National Space Centre. Britain has paid £13 million (US\$21 million) towards the cost of JET-X (Joint European Telescope). But, like other Spectrum X instruments, JET-X has been sitting in storage for several years.

Scientists hope the Moscow meeting will restore confidence that the new launch date of 2001 will be met. Further delays, they fear, may render Spectrum X's scientific capabilities outdated, given that three major X-ray astronomy missions are to be launched by Europe, the United States and Japan in the coming months. **Alison Abbott**

Japan bids to catch up on gene sequencing

[TOKYO] Japan is increasing its effort to develop new technologies in functional genomics, the interpretation of the function of the DNA sequence. It hopes to compete with US and European companies in discovering and patenting human genes.

These projects are in line with the government's recent efforts to commercialize biotechnology research, in particular the new programme aimed at expanding Japan's biotechnology market 25-fold by 2010 (see *Nature* 397, 554; 1999).

But the move also reflects a sense of urgency among researchers that Japan's genome sequencing projects are lagging significantly behind those of the United States and Britain.

The past month has seen a substantial expansion in the number of projects aimed at developing techniques to study DNA function on a genomic scale. These include a joint venture by Tokyo University and 12 Japanese companies to sequence and analyse full-length human complementary DNA (cDNA), in search of genes that will contribute to the understanding of diseases (see *Nature* 398, 644; 1999).

DNA microarray, or 'chip' technology, which allows monitoring of gene expression and detection of mutations, is another area which is rapidly expanding in Japan.

Takara Shuzo, a major distiller of *shochu* (spirits) and a biotechnology company, is planning to commercialize the first domestically produced DNA chip by next March. Leading genome researchers, such as Kenichi Matsubara from the International Institute for Advanced Studies, recently set up a



Wada: budget system prevents fast response.

new venture business based on such technology.

With technological help from the US biotechnology company Genetic Micro Systems, Takara has made chips programmed with cyanobacterium (*Synechococcus*) DNA sequenced by Kazusa DNA Research Centre, an institute dedicated to sequencing and analysing DNA. Other companies, such as Takeda Chemical Industries, are developing DNA chip technology to detect mutations in genes that foster a disposition towards cancer.

These moves take place at a time when Japanese partners in the Human Genome Project, which accounts for five to seven per cent of the total effort, are struggling to meet their target of sequencing 10 per cent of the entire human genome. This has become harder since the US and UK research groups accelerated work in response to the challenge posed by Celera Genomics, the US company set up by geneticist J. Craig Venter.

Japan's limited sequencing capacity was highlighted last month when it was revealed that Celera was planning to sequence the entire rice genome in just six weeks. Japanese researchers were planning to complete the sequencing work in ten years using government funding (see *Nature* 398, 545; 1999).

Japanese researchers are speeding up, for example by improving their sequencing

capability and adopting a new strategy focusing on the gene-rich portion of the genome. But funding for new sequencers will not be available for two years because of the sluggish budget system in Japan.

Akiyoshi Wada, director of the Genomic Science Centre (GSC) — set up last October by the Institute of Physical and Chemical Research — says it would be difficult for Japan to bring forward its sequencing deadline, as Britain's Wellcome Trust and the US National Human Genome Research Institute did with the Human Genome Project.

"One of the main problems is how the current budget system restricts projects from receiving extra funds without going through lengthy and convoluted review processes," says Wada. "The second problem is that research directors do not have enough enforcing power to make important policy decisions — let alone drastic ones."

Many blame Japan's poor track record in genome sequencing on the government's failure to tackle genome research in a coordinated way. In the past, such efforts in Japan have been rather disorderly, with ministries carrying out individual genome projects.

According to Wada, Japan has been effectively excluded from international moves in human genome research, including a decision by the Wellcome Trust and leading international pharmaceutical companies to create a consortium to map single-nucleotide polymorphisms (see *Nature* 398, 545–546). The reason, he says, was Japan's lack of a flagship genome research centre until GSC was launched last year.

"Many researchers have been long aware of the need to build a large centre for genome research, but it took a while to convince both the government and the scientific community to recognize its importance," says Yoshiyuki Sakaki, leader of GSC's human genome research group and a professor at Tokyo University's Institute of Medical Sciences.

GSC focuses on DNA sequencing, functional genomics and the analysis of protein structure and function. It plans to increase its current sequencing capacity of 1,000 megabases per year by three- to fivefold when its new facilities are completed next October.

"We are hoping that Japan will be able to take an active role in post-sequencing research, including the study on variations in human populations," says Sakaki.

Wada emphasizes the significance of a publicly funded genome sequencing effort, and says that GSC will strongly support the policy followed by the UK and US research groups. "The sequence data should be freely and publicly available to the international research community, and we hope to take part in whatever can be done as publicly funded endeavour."

Asako Saegusa

Drug company backs Africa's war on AIDS

[WASHINGTON] A leading maker of AIDS drugs announced last week a major donation to fight the disease in southern Africa. The \$18.3 billion company, Bristol-Myers Squibb (BMS), will give \$100 million over the next five years to a programme of clinical research, physician training and AIDS education, prevention and treatment.

The countries involved are South Africa, Botswana, Namibia, Lesotho and Swaziland, where HIV infection rates in adults are as high as 25 per cent. Most of the work will target women and children. "We feel a moral obligation to take action against this grave situation," says Charles Heimbold, BMS's chairman and chief executive officer.

According to a report in *The Wall Street Journal*, Heimbold began promoting the idea after a dinner-party conversation last year with Kofi Annan, the United Nations secretary general. Some argue that more than altruism is involved: the move comes as

governments in Africa and elsewhere are threatening to ignore patents on expensive AIDS drugs, including several made by BMS.

Later this month, the World Health Assembly, the governing body of the World Health Organisation, will vote on a new policy that may encourage third-world governments to override AIDS drug patents and license home-based companies to make cheap generic versions.

"There's certainly an awareness on the part of companies like Bristol-Myers Squibb that they are vulnerable to criticism on the pricing issue," says Daniel Zingale, the executive director of AIDS Action, a Washington-based lobbying group. Still, he says, the BMS action is "courageous".

But Mark Ahn, BMS's senior director of operations and planning, who is team leader on the Africa project, denies that intellectual property issues play a role. "We felt it was a moral imperative," he says. **Meredith Wadman**

Science advice test in UK devolution vote

Ehsan Masood

With the arrival of devolved governments in Scotland and Wales, British science has an opportunity to build on past experience and experiment with new ways of implementing science advice.

[LONDON] Last week, voters in Scotland and Wales elected representatives to new regional assemblies, ushering in a federal form of government in Britain. The Scottish parliament will be the first in three centuries.

Both the Scottish parliament and the Welsh assembly will be responsible for regional affairs, including health, education and agriculture, and their associated research budgets. The Scottish parliament will be the more powerful of the two, with powers to levy taxes on Scotland's 5 million population.

Scotland and Wales are each expected to be governed by centre-left coalitions of the Labour and Liberal Democrat parties. Parties campaigning for independence from the United Kingdom will form the main opposition in both assemblies.

So far, there has been broad support from both the research community and policy-makers for some of the interim decisions that have been taken which affect science.

There is, for example, overwhelming support for the decision to allow Scotland's scientists to continue to be funded from the main UK research council budget, instead of from a separate, smaller budget administered by a Scotland-only research council.

There has also been praise for Labour's pre-election promises to transfer to Scotland an extra £100 million (US\$163 million) over three years out of the UK science budget, and to appoint a 'science champion' to coordinate science policy across Scotland and to liaise with the UK Office of Science and Technology in London. The Scottish government is also expected to set up a Council for Science and Technology, analogous to that in London, which operates for the whole of Britain.

Despite such promising signs, there is concern that none of the main political parties appears to have a clear policy on three key issues: how the new parliaments will obtain independent advice on scientific issues; whether they will set up committees to monitor their respective governments' handling of science and technology; and what will be the composition of scientific advisory committees attached to government departments.

Research council

Despite vigorous internal debate within the science community and the Scottish Office, the UK government department responsible



Which way next? A devolved government in Scotland will be the UK's first for 300 years.

for Scotland before devolution, the latter's decision not to set up a separate Scottish research council has been largely welcomed — even by the Scottish National Party, the pro-independence main opposition party in Scotland.

University research in Scotland will therefore continue to be funded largely through the existing 'dual-support' system, with the Scottish Higher Education Funding Council paying £120 million each year for basic salaries and infrastructure costs, and the UK research councils paying for research grants.

Scotland's strong research tradition has made it a key target in the UK government's attempts to encourage commercialization of research. It has a strong research base — particularly in food, agriculture, biotechnology and semiconductors.

In addition, Scotland has proportionately more students in higher education than has any other region in the United Kingdom. And its scientists receive more per capita funding than their counterparts in England and Wales.

A separate research council might have led to reduced funds, says Joyce Tait, director of the soon-to-be-launched Scottish Universities Policy, Research and Advisory Network, based at the universities of Edinburgh and Strathclyde.

But the idea of a Scottish research council still has some support among Scotland's five agricultural and biological research institutes, including the Scottish Crop Research Institute near Dundee and the Rowett Research Institute in Aberdeen.

These institutes, which together receive £75 million annually from the Scottish Office, will in future have to lobby the new parliament for their funding allocation. And the parliament, says one scientist, may be tempted to sacrifice research for more immediate concerns, such as education and health.

Science advice

Scotland's academy of science, the Royal Society of Edinburgh, and officials in the Scottish Office, are keen for their new parliament to have an independent source for scientific advice, like the Parliamentary Office of Science and Technology at Westminster.

So far, however, there is little political support for this idea, according to a comparison of the science policies of the five main political parties published last week by the pressure group Save British Science. Its director, Peter Cotgreave, says that science was not high on any party's election agenda.

Cotgreave is among many who believe that devolution provides a rare opportunity for a fresh look at the relationship between science and government after the collapse in public confidence in the UK government's handling of science during the crisis over BSE and, more recently, genetically modified food.

Another such voice is that of Geoffrey Boulton, dean of science and engineering at the University of Edinburgh, who chaired a Royal Society of Edinburgh working group on science and devolution (see *Nature* 396, 402; 1998). Boulton believes that politicians at Westminster have failed to grasp the nature of science, assuming it to be a fixed body of knowledge. He adds that Scotland's politicians will help science — and themselves — by not defending policies purely on the basis of scientific advice; at times, this "is no more than mapping the boundaries of uncertainty".

In many ways, however, the new parliaments promise to be a radical improvement on Westminster. Innovations are to include electronic voting in the chamber, a more consensual style of politics as no party has an overall majority, and more mechanisms for the public to influence legislation.

Donald Bruce, director of the Church of Scotland's Society, Religion and Technology Project, says the effectiveness of pre-legislative scrutiny will be a key test for the new parliaments. He says he is unsure how far they will open up to public consultation.

"The temptation of power is to always hang on to as much as possible, or to open up non-essentials for wider debate," he says. "On the basis of Westminster precedence, I fear that good intentions may not be delivered, once the realities of government begin." □

AP/LOUISA BUTLER

Los Alamos scientist denies leaking weapons secrets

[WASHINGTON] Wen Ho Lee, the scientist accused in US newspapers of leaking nuclear weapons secrets from the Los Alamos National Laboratory, has denied all the accusations. A statement issued by his lawyer in Los Angeles said that he was a "loyal American", and denied the most recent allegation, that classified material found by federal agents on his computer system had been accessed "by any unauthorized person".

Lee said that he and his wife, Sylvia, had worked with the Federal Bureau of Investigation in the 1980s to help gather information about Chinese scientists suspected of spying on the United States. Lee has not yet been charged with any offence, but prosecutors in New Mexico are preparing charges against him.

Sell-off proposal for UK defence research agency

[LONDON] The British Ministry of Defence is considering privatizing part of its Defence Evaluation and Research Agency, according to proposals published last week. But the chemical and biological defence research

laboratories at Porton Down in Wiltshire will be excluded from the sell-off.

Privatization will be controversial. A cross-party select committee of the House of Commons has voiced doubts about the idea because of its potential impact on collaborative research projects. The defence ministry has said that a company could be floated in the financial year 2001/2002. The government would retain ownership of a 'golden share'.

Howard Hughes boosts international grants

[WASHINGTON] The Howard Hughes Medical Institute is to award \$14 million to non-US scientists to study infectious and parasitic diseases. Forty researchers will receive between \$40,000 and \$90,000 a year over five years under the new grants programme, which is open to scientists from all over the world.

It is the first international programme from Hughes — one of the world's largest philanthropic institutions supporting research — to allocate grants to a specific field of research, rather than to a geographical region. Instead of inviting applications direct, Hughes says it is consulting "experts in infectious and parasitic diseases at leading scientific organizations" to obtain nominations for scientists to receive the awards.

Norway could ban the cloning of invertebrates

[OSLO] The Norwegian government is considering legislation to ban the cloning of invertebrates. But the Norwegian Research Council has expressed alarm about the move, which it says would put Norwegian biotechnology research at a huge disadvantage compared with other countries.

The research council's director general, Christian Hombro, has proposed instead a strict system of licensing for working with genetically modified animals. Norwegian legislation already bans human embryo research from the stage of egg fertilization.

Funding plan for Israel's Framework projects

[JERUSALEM] A number of joint Israeli-European research projects that are applying for support from Europe's fifth Framework programme will receive interim funding from the Israeli government.

Orna Berry, chief scientist at Israel's Ministry of Commerce and Industry, says projects that answer to her office's standard criteria for subsidies will receive funds for six to nine months. This will bridge the gap between submission of grant applications to Brussels and approval of Framework funding.

Democrat Brown ready for work after surgery

[WASHINGTON] George Brown (Democrat, California), the senior minority member of the Science Committee in the US House of Representatives, is expected to be back at work this week after undergoing heart valve replacement surgery at the Bethesda Naval Hospital, Maryland.

Brown needed the surgery to correct an impairment that resulted from having rheumatic fever as a child, said a statement from his office. Brown is likely, but not certain, to run for Congress again next year, in the hope that he will regain the chair of the Science Committee if the Democrats regain control of the House.

WHO puts malaria and smoking in its sights

[GENEVA] Action to reduce deaths from tobacco smoking and malaria should be high priorities for the next century, according to the 1999 *World Health Report* published by the World Health Organization this week. The report reiterates its previous call for a global ban on tobacco advertising, and says that deaths from malaria could be halved to 500,000 annually if US\$1 billion more was invested to strengthen healthcare systems.

Malaria accounts for up to one in four child deaths in parts of Africa. The report also warns that the death toll from smoking has been seriously underestimated. "Recent studies suggest that as many as one in two long-term smokers die from their habit."

France wants Euro bank to boost IT lending

[LONDON] France has called on the European Investment Bank to increase lending to technology related companies by 1 billion euros (US\$1.07 billion). The bank is the main long-term lending institution for states within the European Union.

The suggestion was made by Dominique Strauss-Kahn, France's finance minister, after a meeting last weekend with his German counterpart, Hans Eichel. The bank has granted 9 billion euros in loans for information technology (IT) over the past five years.

Biotech not the answer to hunger, says charity

[LONDON] The British charity Christian Aid has challenged biotechnology companies to prove that genetically modified food is the answer to global hunger. In a report, *Selling Suicide: Farming, False Promises and Genetic*

Modification in the Developing World, the charity claims that better food distribution is a more appropriate solution to hunger than improvements in technology.

"The world is not short of food," says the report. "Brazil is the third largest food exporter in the world, and yet 100,000 children die from hunger each year." It says that during the 1984 famine in Ethiopia, "some of the best farming land was used to grow animal feed for export to Europe".

Publisher's announcement

Nature Publishing Group

Macmillan Ltd is today announcing the formation of the Nature Publishing Group. The group will include in its portfolio *Nature* and its related journals, such as *Nature Genetics* and *Nature Cell Biology*, and the 77 journals published or distributed by Stockton Press, including *Oncogene*, *Gene Therapy* and the *European Journal of Cell Biology*.

The Nature Publishing Group's values will reflect those of its progenitor publishers: quality of content, fair pricing and, for authors and learned societies, access to reliable editorial, production and distribution services. For further information about these journals and their publishers, see www.nature.com and www.stocktonpress.co.uk



The full text of items on this page — and further news and background information about the Unesco/ICSU World Conference on Science, to be held in Budapest from 26 June to 1 July — can be found on *Nature's* website at <http://www.nature.com>

Commission seeks to 'blow the whistle' on unethical decisions

[LONDON] The role of Unesco's newly created Commission on the Ethics of Scientific Knowledge and Technology (COMEST) is to "blow the whistle", according to its director-general, Vidgis Finnbogadóttir, speaking after the close of its first meeting in Oslo last month.

A main theme at the meeting was the ethics of the information society. Finnbogadóttir, a former president of Iceland, says it was apparent that very little work had been done in this area, even though "the ethical issues are very clear".

"A greater part of the world does not have access [to information technology], and this alienates them and widens the gap between those who have knowledge and those who do not," she says. "The expense of technology deprives many people — this is unethical."

Finnbogadóttir says that COMEST's mission is to identify problems and find a voice: "We want to define the most burning issues and get the attention of decision-makers."

One key topic raised at the meeting was the issue of irreversible action in the environment, such as the depletion of resources or the destruction of a landscape. Citing a case in Norway in which the building of a hydropower station had created local flooding, Finnbogadóttir said that COMEST would bring such poor ethical decisions to the attention of ordinary citizens.

She emphasizes that problems will be continuing, and that COMEST will give its opinion regularly. It will report on its mission and activities to the World Conference on Science in Budapest.

Full text: <http://helix.nature.com/wcs/a30.html>

US academy backs plan for multinational expert panels

[WASHINGTON] The US National Academy of Sciences is planning to help create an Inter-Academy Centre, run by a multinational board, as a mechanism for setting up and running panels of top-level scientific, engineering and health experts.

Describing the plans in Washington last month, academy president Bruce Alberts said the new body would operate as an international version of the academy's National Research Council. The expert panels would advise global institutions such as the United Nations and the World Bank on issues of critical importance to them.

Alberts said he was keen for the academy to move into the international arena "to ensure that the type of science and technology advice that so wisely informs policy at home can help inform decisions abroad". He pointed out that four years ago it had joined with the scientific academies of other nations to create the InterAcademy Panel on International Issues, an informal network now totalling 80 academies.

It had also developed a website to promote rapid communication between these academies as they prepare for a Conference of Academies, to be held in Tokyo in May 2000, which will address "the many opportunities and challenges for scientists as the world accommodates an estimated 10 billion people in the 21st century".

Alberts pointed out that policy-making institutions will face increasingly complicated issues involving questions of scientific validity and balance. "The world badly needs an impartial mechanism, based only on science, to promote smarter decision-making



Alberts: seeks global solution to problems.

on such issues as agricultural strategies for Africa, safe drinking water in Bangladesh and energy options for Asia," he said.

"The world's academies and their counterpart organizations are the ideal institutions for providing independent, credible, timely, multinational advice

on a broad range of such issues — and we are presently working to help them accept this important responsibility."

In his speech, delivered to the academy's annual meeting, Alberts also urged the world's scientists to work together to create a communication network designed to empower scientists and scientific organizations with valuable knowledge and skills.

He suggested that the world's major scientific organizations cooperate to connect all scientists to the World-Wide Web, where necessary by providing subsidized Internet access through commercial satellite networks, and to take responsibility for generating "scientifically validated knowledge resources", made available free on the web, in preparation for a time when universal Internet access for scientists is achieved in developing and industrialized nations.

"By taking full advantage of new information technologies, the scientific community has an unprecedented opportunity to close the vast 'knowledge gap' between all peoples."

Full text: <http://helix.nature.com/wcs/c15.html>

British government under fire for failing to back Unesco efforts

[LONDON] British supporters of Unesco remain at loggerheads with the government over the extent to which it is prepared to make use of Britain's membership of the UN agency, which it rejoined in 1997 after a 12-year absence.

The Labour government rejoined in order to fulfil a commitment the party had made before coming to power in the 1997 election. Since then, however, Whitehall departments have been reluctant to provide additional financial support for Unesco-related activities in Britain on top of the annual membership fee paid to the agency.

Last week, the two joint chairmen of the

UK Unesco Forum — Dennis Chisman and Malcolm Harper — wrote to Jack Cunningham, the minister for the Cabinet Office, protesting that there appeared to be "overwhelming evidence" that the government's commitment to integrated government was not being followed in the case of policy towards Unesco.

They point out that only one government department — the Department for International Development — is prepared to make a significant contribution to the cost of servicing a national commission.

The Department for Education and Employment had agreed to pay for the work

of a separate committee in its field, but not to help towards common costs. And, while the Foreign and Commonwealth Office had offered a "token contribution" of £5,000 (US\$8,130), other departments, including the departments of trade and industry and of the environment, refuse to contribute.

"The national commissions of all our European Union partners have always been fully funded by government," the two authors say in their letter to Cunningham. "There are real policy issues of policy coherence which urgently call for your involvement."

Full text: <http://helix.nature.com/wcs/b36.html>

Too many agencies spoil Euro research

Sir — Your editorial, “Dangers of Euro-relevance”, urges: “In the absence of strong European alternatives, all the more reason for European governments to keep their national support for basic research” (*Nature* 398, 1; 1999). I disagree with this conclusion which seems to want to maintain the status quo in research funding.

The European system of national research agencies hinders high-quality research. I have just reviewed, for agencies in

the United Kingdom, Switzerland and Spain, grant applications that had significant overlap. As each agency uses a different pool of reviewers, such overlap is unlikely to be spotted normally, so duplicated projects are funded, wasting resources.

Even when the pressure from Republicans to decentralize the US government was at its height, the Democrats won the argument when they asked: “Do you want an Arkansas or a Montana National Institutes of Health?” It

seems that *Nature* advocates just such a development in European science agencies.

The pressure should instead be on the European Union to support basic science, perhaps using the dividend from the reformed Common Agricultural Policy. At a minimum, agencies should draw reviewers from all European countries, instead of only using national lists, as many still do.

Carlo Gambacorti-Passerini

Oncogenic Fusion Proteins Unit, Istituto Nazionale Tumori, Via Venezian 1, 20133 Milano, Italy

Building the future of biocomputing

Sir — It was encouraging to read the News article on the possibility that the US National Institutes of Health might invest more in computing infrastructure (*Nature* 398, 93–94; 1999). But, echoing Larry Smarr’s comment, one wonders whether ‘big iron’ supercomputers are really the way to go.

Many biological calculations increasingly emphasize other aspects of computer hardware than fast central processing units for floating-point calculations. Scientists working on large-scale genome comparisons need lots of storage (on disk and in memory) and high-speed networking perhaps more than ‘mega-flops’.

And those building biological databases are more concerned with the development of common data formats and easy-to-use devices for the bulk entry of laboratory data. One could even argue that these new genome- and database-oriented calculations represent a different paradigm of scientific computing from physically oriented simulations, in need of a distinct but still extensive computational infrastructure.

Mark Gerstein

Molecular Biophysics & Biochemistry Department, Yale University, Bass 432A, 266 Whitney Avenue, New Haven, Connecticut 06520-8114, USA

Phenology and the changing seasons

Sir — Phenological data (observing seasonal plant and animal activity driven by environmental factors) have recently been used to suggest that the length of the growing season in Europe has increased over the past 30 years by perhaps a week in both spring and autumn¹. These data could be interpreted

somewhat differently, particularly those relating to the onset of autumn.

Clearly, the date of bud burst is a reliable indicator of the earliness of spring. To break quiescence, buds must accumulate a particular heat sum, in degree-days, above a certain threshold temperature. This is usually cited as 5 °C, presumably the point below which respiration is negligible. Climate changes occurring more than 1,000 years ago have been inferred from the recorded dates of such phenology-related events as the Kyoto cherry blossom festival². Rarely, although perhaps increasingly frequently if the climate is getting warmer, there may be insufficient winter chilling to break rest completely, so that a late bud-burst date could mean a mild winter, not a cool spring.

Choosing a reliable phenological event with which to judge the arrival of autumn is more problematic, and dates of leaf colour change and leaf drop may be inappropriate. Various factors have been cited as causing these events, including cold, drought, photoperiod, clear weather and injury³. But leaves age⁴. Just as buds have a required heat accumulation for flushing, it seems that leaf life and senescence are also mediated by heat accumulation, at least in some species⁵. This is more likely in determinate species such as oaks, where all leaves are roughly of one age cohort². A later end to the growing season, as judged by later leaf fall/drop dates, could be due to the fact that the leaves’ heat sum is not met as early as in previous years because of later bud burst and/or cool summer temperatures, rather than a mild autumn. Incorporating summer temperatures into the model might usefully modify the conclusions about the date of the end of the growing season.

More reliable ways of assessing the effects of environment on phenology require more accurate estimates of phenological factors. The real start of spring is when mitosis starts in tissues adjacent to shoot apical meristems; the end of the growing season is when mitosis stops in the apical meristem. Unfortunately, these

events are far more difficult to observe, and require destructive sampling.

John Worrall

Forest Sciences, University of British Columbia, 2424 Main Mall, Vancouver V6T 1Z4, Canada

1. Menzel, A. & Fabian, P. *Nature* 397, 659 (1999).
2. Lamb, H. H. *Climate Past, Present and Future* (Methuen, London, 1977).
3. Kozlowski, T. T., Kramer, P. J. & Pallardy, S. J. *The Physiological Ecology of Woody Plants* (Academic, New York, 1991).
4. Noodén, L. D. & Leopold, A. C. (eds) *Senescence and Aging in Plants* (Academic, New York, 1988).
5. Worrall, J. J. *Sust. Forestry* 1, 1–18 (1993).

The right prescription for preclinical teaching

Sir — The critical and trenchant letter by the pharmacologists Hartmut Glossmann and Bernhard Peskar on the status of Austrian medical training is welcome (*Nature* 396, 614; 1998). But, as medical doctors, members of the teaching staff and students of the Medical School in Graz, we take issue with the statement that “Many teaching staff lack an MD qualification and do not know the basic requirements of clinical work”. This statement is bewildering because, in all clinical branches, members of the teaching staff must have an MD qualification, and do know the basic requirements of clinical work.

Presumably, Glossmann and Peskar meant that in many preclinical branches — biochemistry, pharmacology and physiology to name but a few — teachers of medical students do not always have an MD but instead a PhD qualification. So, the question arises whether effective and medically oriented teaching of medical students in these preclinical disciplines requires an MD qualification. In our estimation, it is not a prerequisite at all.

H. Peter Soyer, Birger Kränke, Thomas Petnehazy, Stefan Scheidl

Department of Dermatology, University of Graz, Auenbruggerplatz 8, A-8036 Graz, Austria

Dinosaur tracks in the computer age

Kevin Padian

A three-dimensional record of dinosaur feet and movement comes from 200-million-year-old footprints made in wet mud. Comparisons of these prints with the tracks made by living birds clear up some of the mysteries about dinosaur toes and the tracks that they left.

It was no accident that Sherlock Holmes had the habit of rousing his faithful companion off to another adventure with the exhortation, "Come, Watson! The game is afoot". Their creator, Arthur Conan Doyle, was fond of footprints — even fossil ones — and he knew how they could be useful forensically¹. Alas for the many palaeontologists who, through the years, have looked down their noses at such tracks and traces. For footprints have revealed many secrets of locomotion, ecology and behaviour to those who have been patient and sharp enough to study them²⁻⁴. And on page 141 of this issue, Stephen M. Gatesy and his colleagues⁵ bring the science of fossil trackways into the twenty-first century, in ways that Conan Doyle and his creations could never have dreamed.

The story begins as the researchers explored the tree-barren fields of East Greenland. Lured to the Triassic (over 200-million-year-old) exposures near the Fleming Fjord Formation by the prospect of dis-

covering early mammals and their relatives, Gatesy *et al.* found a surprisingly diverse fossil fauna, the components of which are still being described⁶. As well as the bones and teeth of various vertebrates, the authors uncovered strange trackways with features that would have been ignored by most workers because they are so indistinct. Instead, Gatesy and colleagues turned the find into a model for future work.

Baird's 'First Law' of ichnology — the science of footprints — states that a trackway is not a simple record of anatomy. Instead, it is a record of how a foot behaves under a particular locomotory pattern as it makes contact with a particular substrate⁷. The varying conditions of the substrate can have a substantial effect on the features of the trackway, as anyone who has walked along a beach, both close to and above the strand line, can tell — and aching calf muscles, after a good walk along the shore, attest to the influence of substrate on locomotion.

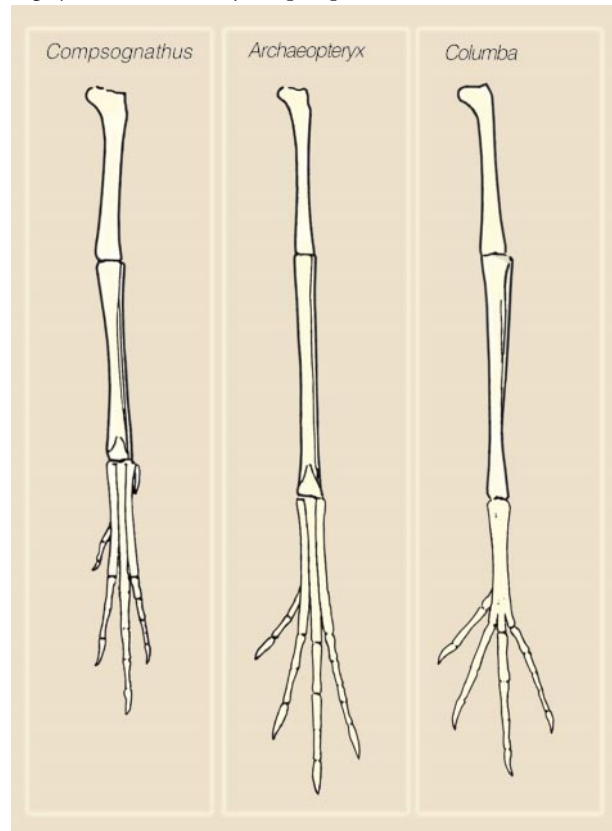


Figure 1 Putting your foot in it — feet from a theropod dinosaur (*Compsognathus*), the first known bird *Archaeopteryx*, and a pigeon (*Columba*). Gatesy *et al.*⁵ studied the footprints made by theropods in wet mud, and worked out how the feet of these dinosaurs compare with those of living birds. Their results show that living birds walk much like their ancestors did, although the dinosaurs carried their heels just a bit lower. (Adapted from ref. 12.)

The tracks studied by Gatesy and colleagues ranged from clear imprints to virtually indistinct traces, depending on the condition of the substrate. They were made by a theropod (carnivorous) dinosaur in mud that was often so sloppy that there was little chance of preserving precise records of individual joints or skin impressions. But this sloppiness preserved the entry and exit 'wounds' made by the foot, which led to an interesting discovery — the deeper you sink, the more of the movement that normally occurs above ground level can take place below it instead.

In theropod dinosaurs, the fifth toe is completely reduced and lost. The first toe (hallux) is short, and it is suspended from the second metatarsal (sole bone), a bit less than halfway up the sole. In those theropods that are closer to birds, the first toe has descended in the course of evolution, eventually hanging from near the end of the sole. For instance, in *Archaeopteryx*, which is the first known bird, the first toe is fully opposable (that is, it faces the other digits on the same foot), and in later birds the claws enlarge for perching, suggesting the start of true arboreality⁸. So, the distinction between bird and other dinosaur tracks has sometimes been assessed on the basis of the imprint of the hallux — if it extends backwards and towards the midline, the maker of the track is often regarded as avian⁹.

In the Greenland tracks, the hallux seems to make such an impression going in — in apparently avian fashion — but on the way out it disappears. To find out why, Gatesy *et al.*⁵ sectioned the fossilized footprints and traced the disappearance to the fact that the toes were brought together as the animal lifted its foot. They then ran guinea fowl and turkey through successively more sloppy mud to demonstrate that this kinematic pattern is simply inherited by today's birds from their theropod ancestors¹⁰. Moreover, the footprints elongate as the mud becomes sloppier. In Mesozoic Era trackways, this feature has sometimes promoted the inference that some dinosaurs were plantigrade (that is, they walked on their soles)^{2,3}. But the sloppiness of the sediment now reveals that, in these dinosaurs, the heel was just carried a bit lower than in birds today (Fig. 1). This finding indicates that the stride was more strongly powered by the femur in basal theropods than in birds, where the femur is more stable and the lower leg and foot provide more of the power thrust.

The results of Gatesy and colleagues' investigation are dramatically shown by computer graphics (Fig. 3 of their paper on page 143), which graft the anatomy of a typical basal theropod foot onto the footfall pattern of living birds, allowing for differences in proportions and kinematics. The conclusions are clear — early Mesozoic theropods walked much, but not exactly, like their

living avian descendants. And, more importantly, locomotion and limb function have evolved like any other features¹⁰.

Most of the fossil footprint literature documents new tracksites, describes the form and proportions of tracks, and tries to assign such tracks to trackmakers, usually with little in the way of direct anatomical reference². At a landmark conference¹¹ in 1985, there was consensus that two frontiers should receive renewed attention: kinematic patterns and 'competency' of the sediment. Unfortunately, few studies have since done so. But Gatesy *et al.*⁵ set the standard for future work, and show just how much we have to gain from such analyses. □

Kevin Padian is in the Department of Integrative Biology and Museum of Paleontology, University of California at Berkeley, California 94720-3140, USA.

e-mail: kpadian@socrates.berkeley.edu

1. Batory, D. & Sarjeant, W. A. S. *J. Arthur Conan Doyle Soc.* **8**, 89–101 (1998).
2. Lockley, M. G. *Tracking Dinosaurs: A New Look at an Ancient World* (Cambridge Univ. Press, 1991).
3. Thulborn, T. *Dinosaur Tracks* (Chapman & Hall, London, 1990).
4. Murie, O. J. *A Field Guide to Animal Tracks* (Houghton Mifflin, Boston, 1968).
5. Gatesy, S. M., Middleton, K. M., Jenkins, F. A. Jr & Shubin, N. H. *Nature* **399**, 141–144 (1999).
6. Jenkins, F. A. Jr *et al. Meddr. Grönland, Geoscience* **32**, 1–25 (1994).
7. Baird, D. T. *Bull. Mus. Comp. Zool.* **117**, 449–520 (1957).
8. Padian, K. & Chiappe, L. M. *Biol. Rev.* **73**, 1–42 (1998).
9. Lockley, M. G., Yang, S. Y., Matsukawa, M., Fleming, F. & Lim, S. K. *Phil. Trans. R. Soc. Lond.* **336**, 113–134 (1992).
10. Padian, K. & Olsen, P. E. in *Dinosaur Tracks and Traces* (eds Gillette, D. D. & Lockley, M. G.) 231–241 (Cambridge Univ. Press, 1989).
11. Gillette, D. D. & Lockley, M. G. (eds) *Dinosaur Tracks and Traces* (Cambridge Univ. Press, 1989).
12. Ostrom, J. H. *Annu. Rev. Earth Planet. Sci.* **3**, 55–75 (1975).

Earth science

Radon and rock deformation

Evelyn Roeloffs

What happens when stress is applied to rocks in the Earth's crust so that the crust deforms? This is a question tackled by Trique *et al.* on page 137 of this issue¹. They have used a natural laboratory in the French Alps — the Roselend reservoir — to monitor the geophysical signals that result from the greater or lesser pressure on the underlying crust exerted by the weight of water in the reservoir. This area is not itself prone to earthquakes. But the broader interest of this work is in what it may tell us about the events, induced by crustal deformation, that precede earthquakes.

The ability to predict earthquakes is of course highly desirable. But progress in this difficult and highly contentious science will depend on detecting and interpreting physical changes stemming from the processes

of earthquake generation. Many possible precursors have been reported, but seismologists are sceptical of those that are not clearly linked to crustal deformation. This 'unproven' category includes the well-documented precursory decrease and increase of radon concentration before the 1978 Izu–Oshima earthquake in Japan² (Fig. 1), as well as the controversial assertion that

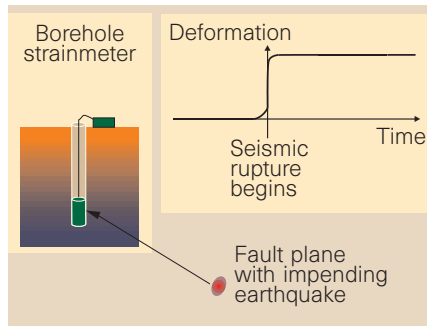


Figure 2 Rock friction, which depends on slip rate and sliding-induced changes on a fault surface, implies that seismic slip should be preceded by accelerating aseismic slip near the hypocentre of an impending earthquake. Sufficient aseismic slip would produce near-surface deformation detectable by a borehole strainmeter. Compared with the strain step recorded at the time of the earthquake, the precursory strain signal would be in the same direction but of much smaller amplitude. A magnitude-5 earthquake, 10 km deep, produces maximum near-surface strain of about 10^{-7} at a site 5 km from its fault plane; strain increases 30-fold for each unit increase of magnitude, but falls off as the third power of distance from the source. Estimates of pre-seismic slip duration and amplitude range widely because frictional parameters of natural faults are poorly known.

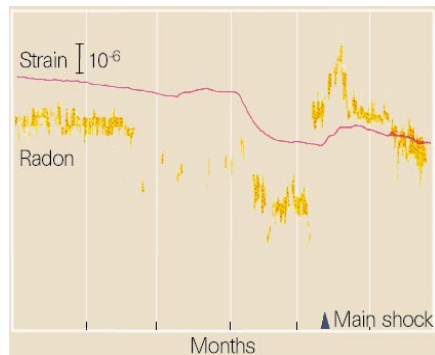


Figure 1 The radon and strain data for the magnitude-7 Izu–Oshima earthquake^{2,9} of 14 January 1978 show changes preceding the earthquake. But they do not match the model shown in Fig. 2; in particular, neither change is monotonic, and in both cases the pre-earthquake change exceeds that produced by the earthquake itself.

moderate earthquakes in Greece have been predicted from variations in the local electric field³.

Trique *et al.*¹ now report that various phenomena — bursts of radon gas, changes of electric potential, and departures of ground tilt from that predicted on an assumption of linear rock elasticity — consistently accompany water-level variations behind the Roselend dam. These phenomena occur together, usually within days after an abrupt change in the reservoir's filling or emptying rate. At Roselend, radon emissions and electrical changes are produced by a well-quantified driving force — the lake 'load', or weight of water — instead of the poorly understood processes that precede earthquakes. But the observations do provide indications of the relationship between radon and electrical anomalies, and the deformation of crustal rocks.

Seismologists expect earthquake precursors to take the form of transient crustal-strain signals from 'aseismic' fault slip near the earthquake's nucleation point (that is, fault slip that is too slow to radiate seismic waves) (Fig. 2). Numerical simulations show, however, that such signals would be exceedingly small⁴. Even the best existing instruments — borehole strainmeters with resolution exceeding a part per billion — would need to be within a few kilometres of the impending earthquake's epicentre to detect this aseismic strain. Although strain changes preceding two California earthquakes have been identified^{5,6}, they don't resemble the expected signals.

Proponents of earthquake prediction maintain that changes in radon emission, or in electrical or magnetic fields, represent a natural amplification of pre-earthquake deformation under special geological conditions. For example, the conductance by rock fractures of water or gas is proportional to the third power of the fracture's aperture⁷. Fluid flow past ions adsorbed on rock surfaces produces an electric field, termed a 'streaming potential', that varies with pressure gradient and permeability⁸. Fluid, gas or electromagnetic measurements might thus detect deformation indirectly, albeit at localized sites and with amplitudes related nonlinearly to strain.

Silver and Wakita⁹ list many potential examples of such pre-earthquake 'strain indicators'. Unfortunately, these indicators are irreproducible: they can be detected only in certain locations, but in any one location earthquakes recur infrequently. What is needed is evidence that transient strain leads consistently, if not linearly or uniformly, to observable phenomena. The radon, electrical and ground-tilt measurements from Roselend lake constitute this kind of reproducible evidence.

The shallow crust's reaction to large changes in lake level may also illuminate the

physics of earthquake generation. Although Roselend lake is free of earthquakes, the apparent stimulation of geophysical anomalies by changes in its filling or emptying rate is reminiscent of peculiar seismic features evident at other large reservoirs. Earthquakes larger than magnitude five at Koyna reservoir (India) are induced when the lake level rises faster than 12.2 m per week¹⁰ and, at Nurek reservoir (Tadjikistan), the largest local earthquakes occurred during abrupt decreases in the reservoir filling rate¹¹. So the Roselend observations might provide clues as to how, at Koyna and Nurek, loading rate or fluid-flow rate influence the behaviour of brittle, inelastic rock.

How do crustal rocks respond inelastically to loading by a surface reservoir? Although induced stresses and pore pressures vary spatially around a reservoir, they are unlikely to be much greater than 0.1 MPa for each 10 m of water depth. Nonlinear behaviour of rock mass seems to occur at low stress, and must therefore be dominated by deformation of fluid-filled void space, including cracks and faults. One question is whether stress concentration at crack tips allows the creation of new crack surfaces under loads such as those imposed by lakes, which could increase permeability, facilitate radon dissipation and locally weaken the rock mass. Another issue is whether the loading-rate dependence of rock-mass deformation reflects the frictional character of fractures or time delays due to pore-fluid diffusion. Trique and colleagues¹ attribute radon and electric-field anomalies to changing rates of fluid flow in the rock, a

hypothesis that could be tested by monitoring groundwater flow near Roselend lake.

More than 20 years ago, earthquake scientists demonstrated that raising and lowering subsurface fluid pressure could turn microearthquakes on and off in Colorado's Rangely Oil Field¹². Laboratory experiments had already shown that increased pore-fluid pressure counteracts the compressive stress on fault planes in rock, reducing the shear stress required for slip, but it was the Rangely results that really caught seismologists' attention. Large-scale field experiments, such as those carried out by Trique *et al.*, are orders of magnitude closer in size to that of earthquake-producing faults, and they encompass the natural fractures that concentrate deformation and conduct fluid. Work at this scale can determine how radon emission and electric fields are affected by crustal deformation, fluid flow and specific geological structures. And that might clarify their relevance to earthquake prediction. □

Evelyn Roeloffs is at the US Geological Survey, 5400 MacArthur Boulevard, Vancouver, Washington 98661, USA.

e-mail: evelynr@usgs.gov

1. Trique, M., Richon, P., Perrier, F., Avouac, J. P. & Sabroux, J. C. *Nature* **399**, 137–141 (1999).
2. Wakita, H., Nakamura, Y., Notsu, K., Noguchi, M. & Asada, T. *Science* **207**, 882–883 (1980).
3. Varotsos, P., Alexopoulos, K., Lazaridou-Varotsou, M. & Nagao, T. *Tectonophysics* **224**, 269–288 (1993).
4. Stuart, W. D. & Tullis, T. E. *J. Geophys. Res.* **100**, 24079–24099 (1995).
5. Gladwin, M. T., Gwyther, R. L., Higbie, J. W. & Hart, R. H. G. *Geophys. Res. Lett.* **18**, 1377–1380 (1991).
6. Roeloffs, E. & Quilty, E. *Pure Appl. Geophys.* **149**, 21–60 (1997).
7. Committee on Fracture Characterization and Fluid Flow *Rock Fractures and Fluid Flow: Contemporary Understanding and Applications* (National Acad. Press, Washington DC, 1996).
8. Jouriaux, L. & Pozzi, J.-P. *J. Geophys. Res.* **100**, 10197–10209 (1995).
9. Silver, P. G. & Wakita, H. *Science* **273**, 77–78 (1996).
10. Gupta, H. K. *Bull. Seismol. Soc. Am.* **73**, 679–682 (1983).
11. Simpson, D. W. & Negmatullaev, S. K. *Bull. Seismol. Soc. Am.* **71**, 1561–1586 (1981).
12. Raleigh, C. B., Healy, J. H. & Bredehoeft, J. D. *Science* **191**, 1230–1237 (1976).

Cell signalling

Calmodulin at the channel gate

Michael D. Ehlers and George J. Augustine

Of all the channels that conduct ions in response to voltage changes across cell membranes, those that carry calcium merit special attention. The Ca^{2+} that enters through these channels acts as a messenger for a host of intracellular signalling events, including feedback

processes that regulate activity of the channel itself. Such feedback includes inactivation¹, which closes the channel, and facilitation², which enhances channel opening. A number of studies^{3–6}, including reports by Lee *et al.*³ and Zühlke *et al.*⁴ (pages 155 and 159 of this issue), have now reached the surprising conclusion that calmodulin, which is the classical Ca^{2+} receptor protein inside cells⁷, mediates both inactivation^{3–6} and facilitation^{3,4}.

Most of the new studies focused on the L-type Ca^{2+} channel, which is involved in contraction of heart muscle, gene expression in neurons, and hormone secretion. The first hint that calmodulin might be involved in Ca^{2+} -dependent regulation of these channels was the observation that deleting part of the channel's cytoplasmic tail prevents inactivation⁸. This deleted region includes a sequence of amino acids that resembles an isoleucine–glutamine (IQ) motif, a domain which, in many other proteins, binds calmodulin.

Four new findings support the idea that calmodulin regulates the properties of Ca^{2+} channels, and that it does so by binding to the IQ-like motif. First, Zühlke *et al.*⁴ and Peterson *et al.*⁶ show that mutant forms of calmodulin, which cannot bind Ca^{2+} , act as dominant-negative inhibitors of Ca^{2+} -channel inactivation. Although this finding establishes that calmodulin helps to inactivate Ca^{2+} channels, it does not indicate where calmodulin acts. Calmodulin could, for example, bind directly to the channel, or it may have other targets, such as protein

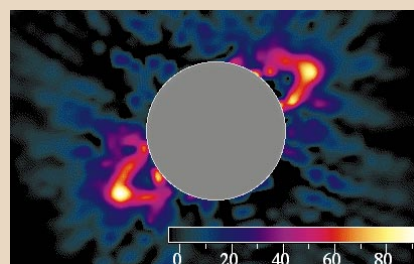
Astrophysics

The planet factory

The first near-infrared image of a dust ring around a young, nearby star has been taken by the Near Infrared Camera and Multi-Object Spectrometer (NICMOS) on board the Hubble Space Telescope. The image may give astronomer Glenn Schneider and his colleagues a new look at early planet formation (*Astrophys. J.* **513**, L127–L130; 1999).

Planets are thought to condense out of the disks of dust and gas surrounding newborn stars by a process of accretion—in which small particles collide and stick together. Such circumstellar disks are hard to see, because the glare from the central star outshines the weaker, reflected light from the disk. This particular image was captured using a coronagraphic camera on NICMOS to block out the glare of the star (grey circle on the image).

This disk is unusual as it is only 1.6 billion miles wide (the ring diameter is 13 billion miles), leaving a large dust-free area inside the ring. The much smaller rings found around planets such as Saturn are



held in place by the gravitational force of moons orbiting nearby. The narrowness of this stellar ring implies that it may be confined by one or more unseen bodies—probably new planets. Without some mechanism to keep them intact, dust rings around stars would spread outwards, reducing their ability to form planets.

The colour of the ring is somewhat reddish (scale bar gives flux density in μJy per pixel), showing that it is made up of grains several μm in size, which is larger than typical interstellar dust. This stellar ring is surprisingly young, indicating that planets may have formed in less than ten million years.

Sarah Tomlin

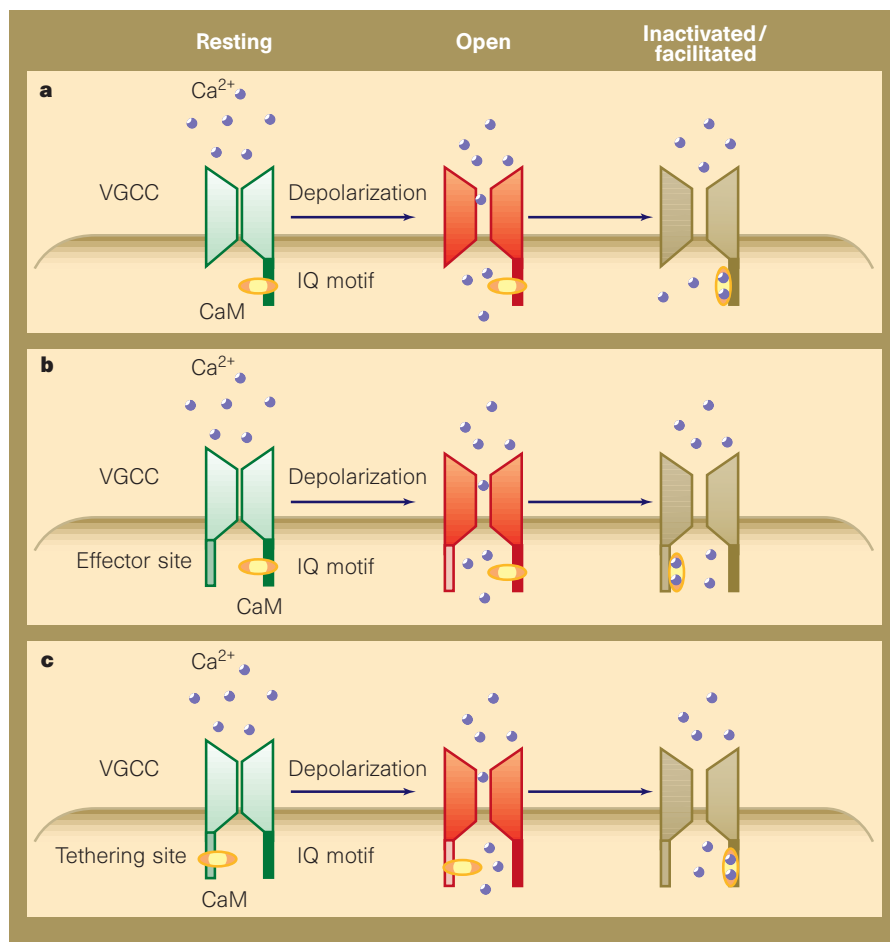


Figure 1 Three models for calmodulin-dependent regulation of voltage-gated calcium channels (VGCCs), based on the results of Lee *et al.*³, Zühlke *et al.*⁴ and others^{5,6}. In each model there is a tethering site, where calmodulin is constitutively bound under resting conditions, and a calmodulin-effector site through which Ca²⁺-associated calmodulin modulates channel gating. Once the channel is activated by membrane depolarization, tethered calmodulin is activated by incoming Ca²⁺. The calmodulin then binds to an effector site, which closes (inactivates) or further opens (facilitates) the channel. The calmodulin-binding isoleucine–glutamine (IQ) motif on the cytoplasmic tail of the channel may act: a, as both the tether and the effector; b, as the tethering site only; or c, as the effector site only.

kinases or phosphatases, that regulate Ca²⁺ channels⁹.

Second, several biochemical strategies^{4–6} have shown that calmodulin binds to the IQ motif of the L-type Ca²⁺ channel. Binding requires sub-micromolar concentrations of Ca²⁺ — similar to those needed to inactivate Ca²⁺ channels¹⁰ — making the IQ motif a good candidate for the site at which calmodulin acts on the channel. Third, mutations that decrease calmodulin binding to the IQ motif reduce or eliminate the Ca²⁺-dependent inactivation of L-type channels^{4,5}. In the most dramatic case, Zühlke *et al.* found that inactivation was greatly slowed when they replaced the isoleucine (I) of the IQ motif with glutamate, a substitution that markedly reduces calmodulin binding.

Finally, when Zühlke *et al.* replaced the same isoleucine residue with alanine, they found that the L-type channels were facilitated upon repeated activation. Similar alanine substitutions reportedly have little

effect on calmodulin binding⁵, but, as with calmodulin-dependent channel inactivation, these authors found that facilitation is blocked by Ca²⁺-insensitive calmodulin mutants and abolished by replacing the isoleucine in the IQ motif with glutamate. These results indicate that the action of calmodulin on the IQ motif can also facilitate L-type Ca²⁺ channels.

The exciting picture to emerge, then, is that Ca²⁺ both inactivates and facilitates Ca²⁺ channels by binding to calmodulin, and that both processes involve binding of Ca²⁺-calmodulin to the IQ motif. Because Ca²⁺-insensitive calmodulin mutants prevent incoming Ca²⁺ from inactivating or facilitating the channel^{4,6}, calmodulin is probably tethered to (or near) the channel before Ca²⁺ enters. This agrees with earlier studies¹¹, showing the Ca²⁺-dependent inactivation of Ca²⁺ channels reconstituted in artificial membranes (which lack cytoplasmic proteins)¹¹. The close spatial proximity of calmodulin and the channel could also

explain why inactivation is resistant to treatments that act only on freely diffusible Ca²⁺ (ref. 12) or on unbound calmodulin¹³.

Because L-type Ca²⁺ channels are neither inactivated nor facilitated under resting conditions, they must have both a tethering site, which anchors the calmodulin, and a calmodulin-effector site, which gates the channel (Fig. 1). The IQ motif could serve as tethering site, effector site, or both. Zühlke *et al.*⁴ show that some calmodulin can bind to this motif at the concentrations of Ca²⁺ found in resting cells, so perhaps calmodulin is pre-bound to the IQ motif. For the IQ motif to function as both the tethering and effector sites, an influx of Ca²⁺ must cause the calmodulin to undergo a conformational change yet remain bound to the IQ motif (Fig. 1a). Alternatively, the Ca²⁺ influx might cause calmodulin to bind, or bridge, a separate effector site (Fig. 1b). Finally, calmodulin might be tethered to another site near the channel pore and, on binding Ca²⁺, rapidly translocate to the IQ motif (Fig. 1c)⁶.

Calmodulin's sphere of influence even extends to the P/Q-type channels — neuronal Ca²⁺ channels that are not thought to be regulated by Ca²⁺. Lee *et al.*³ and Peterson *et al.*⁶ show that calmodulin binds, in a Ca²⁺-dependent manner, to two different sites in the cytoplasmic tail of the P/Q-type channel. One site, which is homologous to the IQ motif found in L-type Ca²⁺ channels⁶, has also been detected in yet another Ca²⁺ channel, the R-type channel⁶. The second site is not an IQ motif, and is found downstream of the first³. These results indicate that P/Q-type channels — and, perhaps, even other Ca²⁺ channels — may have not one but two calmodulin-binding sites. If so, it will be important to find out whether the different calmodulin-binding domains have different functions. Much like the case for L-type channels, Lee *et al.* show that binding of calmodulin to the second domain imparts a modest Ca²⁺-dependence to inactivation, and is required for long-lasting Ca²⁺-dependent facilitation. As yet, the function of the first, IQ-like, motif of P/Q-type channels is not known.

The new observations mean that voltage-gated Ca²⁺ channels enter the company of other ion channels — ranging from Ca²⁺-activated potassium channels¹⁴ to NMDA (N-methyl-D-aspartate) receptor channels¹⁵ — whose activity is regulated by calmodulin binding. By illuminating the molecular events that regulate the function of Ca²⁺ channels, these papers bring us one step closer to understanding how cells encode Ca²⁺ signals. Future studies must determine whether the IQ motif alone acts as both a tethering and an effector site, or whether additional calmodulin-binding sites are involved. We will then be in a better position to comprehend how just one molecule,



100 YEARS AGO

The possessors of certain hereditary characters are unquestionably **sub-prolific**; that is, they eventually contribute less than their average share to the stock of the future population. It may be that they die before the age of marriage, or that they are sexually unattractive or unattracted, or that if married they are comparatively infertile, or that if married and fertile the children are too weakly to live and become parents. It is very probable, though I have no trustworthy facts to confirm the belief, that persons affected with hereditary insanity are sub-prolific because their families, if they have any, are apt to contain members who are afflicted in various ways that render them less likely than others to live and to marry. But I do not propose to go into the details of this or of any other malady, but merely mention it as an illustration of what is meant, when I assume that the possessors of some particular characteristic, not necessarily a morbid one, and which may be called A, are sub-prolific on the average.

From *Nature* 11 May 1899.

50 YEARS AGO

Reports of tumours in insects are relatively rare. Paillot mentions proliferation in the fat cells of *Euxoa segetum* Schiff., following infection by virus diseases (pseudo-grasseries I and II). Tumours have been described in the fruit-fly, *Drosophila melanogaster* Meigen, by Stark and Russell: in larvae of the Pygaera group of butterflies, by Federley; and in the stick insect, *Dixippus morosus* Br., by Pflugfelder. They have been found in a large Orthopteran insect, *Leucophaea maderae* F., by Scharrer. All the tumours so far discovered in insects are apparently non-malignant, although malignancy has been claimed by Stark and Federley. Russell found that those in *Drosophila* reported by Stark as malignant were similar in structure to benign tumours occurring in the same insect, and that the so-called malignant tumours could be successively transplanted without hampering the development of the host. The tumour strains discovered by Federley apparently have been lost. Various stimuli will provoke tumour proliferations in insects. Spontaneous tumours occur in several genetic tumour strains in the fruit-fly, but the stimulus is not known.

From *Nature* 14 May 1949.

calmodulin, can both inactivate and facilitate Ca^{2+} channels. □

Michael D. Ehlers and George J. Augustine are in the Department of Neurobiology, Duke University Medical Center, Box 3209, Durham, North Carolina 27710, USA.

e-mails: ehlers@neuro.duke.edu

georgea@neuro.duke.edu

1. Eckert, R. & Chad, J. E. *Prog. Biophys. Mol. Biol.* **44**, 215–267 (1984).
2. Noble, S. & Shimoni, Y. *J. Physiol. (Lond.)* **310**, 57–75 (1981).
3. Lee, A. et al. *Nature* **399**, 155–159 (1999).
4. Zühlke, R. D., Pitt, G. S., Deisseroth, K., Tsien, R. W. & Reuter, H. *Nature* **399**, 159–162 (1999).

5. Qin, N., Olcese, R., Bransby, M., Lin, T. & Birnbaumer, L. *Proc. Natl Acad. Sci. USA* **96**, 2435–2438 (1999).
6. Peterson, B. T., DeMaria, C. D., Adelman, J. P. & Yue, D. T. *Neuron* **22**, 549–558 (1999).
7. Means, A. R. *Recent Prog. Hormone Res.* **44**, 223–262 (1998).
8. Zühlke, R. D. & Reuter, H. *Proc. Natl Acad. Sci. USA* **95**, 3287–3294 (1998).
9. Armstrong, D. & Eckert, R. *Proc. Natl Acad. Sci. USA* **84**, 2518–2522 (1987).
10. Johnson, B. D. & Byerly, L. *J. Physiol. (Lond.)* **462**, 321–347 (1993).
11. Haack, J. A. & Rosenberg, R. L. *Biophys. J.* **66**, 1051–1060 (1994).
12. Neely, A., Olcese, R., Wei, X., Birnbaumer, L. & Stefani, E. *Biophys. J.* **66**, 1895–1903 (1994).
13. Imre, J. P. & Yue, D. T. *Neuron* **12**, 1301–1318 (1994).
14. Xia, X.-M. et al. *Nature* **395**, 503–507 (1998).
15. Ehlers, M. D., Zhang, S., Bernhardt, J. P. & Hagan, R. L. *Cell* **84**, 745–755 (1996).

Materials science

The hard problem of carbonitrides

Barry M. Klein

‘Materials by design’ is the goal of theorists who are developing a fundamental understanding of the macroscopic properties of materials from a microscopic, atomistic point of view. In a paper on page 132 of this issue, Jhi et al.¹ move us towards this goal by relating the hardness properties of carbonitrides — compounds containing carbon or nitrogen, together with a transition metal² — to fundamental aspects of the material’s electronic structure such as the electronic energy bands, or the charge density and chemical bonding. They show that the elastic properties of the carbonitrides, which relate to their hardness, can be explained by the filling of electron states depending on whether the underlying lattice is occupied by transition-metal or carbon and nitrogen atoms, or by vacancies. The authors have used state-of-the-art theoretical and computational methods to reach this conclusion from first principles, without recourse to tweaking the theoretical parameters with experimental data. An understanding of the hardness of the carbonitrides does not seem so hard any longer.

Carbonitrides (such as TiC or NbN) often form in the simple NaCl crystal-lattice structure of ordinary table salt. The valence electron concentration, or number of electrons that are active in the bonding of these materials (there are eight such electrons in TiC, and ten in NbN) can be changed by forming alloy compounds with varying amounts of transition metal(s), carbon and nitrogen atoms, or vacancies. Such materials often have interesting properties ranging from superconductivity at relatively high temperatures^{3,4} (for example, $T_c > 17$ K for NbN), to high melting temperatures, and to the extreme hardness (such as TiC, which has a relative hardness somewhere between that of aluminium oxide and diamond) that makes them important materials for cutting tools and related applications.

Hardness is defined empirically as resistance of a material to denting or scratching. Relative hardness may be determined by a material’s response to an ‘indenter’ pressed into, or moved along, its surface. The hardness and related mechanical properties of materials often involve complicated phenomena such as the motion or pinning of dislocations (mismatched lattice planes), or other ‘large defect’ properties of solids that, ultimately, should relate back to the fundamental microscopic properties of the perfect lattice. Theoreticians have long nibbled away at the problem of relating the hardness of carbonitrides to the underlying chemical bonding that is governed by the electron distribution in these materials. But unravelling the underlying properties of carbonitrides in terms of quantifiable physical and chemical quantities such as ‘bonding’ has been limited to qualitative descriptions of these materials, such as describing them as ‘covalent metals’, in analogy to hard semiconductor materials such as diamond.

It is only in the past decade that theoreticians have been able to study electronic structures with enough accuracy to be able to determine fundamental properties such as elastic behaviour and lattice-vibrational properties of solids from first principles. Using the formalism developed by Walter Kohn and colleagues (see for example, ref. 5) that led to his share of the 1998 Nobel Prize in Chemistry, an elaborate but computationally workable machinery for doing quantitative studies of the electronic structure of materials has become possible (see ref. 6 and references therein). In essence, the resulting so-called Kohn–Sham equations reduce the solid-state many-body problem (roughly 10^{23} electrons and nuclei for each cubic centimetre of a solid) to a set of one-particle Schrödinger equations that need only be solved in a single unit cell of a perfect, periodically repeating solid. The results for properties of solids in the ground state

— such as structural features (bond lengths), elastic behaviour and lattice vibration frequencies — has in many cases agreed with experiment to within a few per cent. The only experimental input into these calculations is an assumed crystal structure, but even the ground-state crystal structure can be determined by these methods using a total-energy-minimum search procedure⁷.

The observed maximum in hardness of carbonitrides occurs for a non-integral value of the valence electron concentration, corresponding to major deviations from the properly occupied NaCl lattice structure. In general, such a loss of lattice periodicity would render accurate theoretical calculations impossible. However, one of the remarkable properties of the carbonitrides is their 'rigid band' behaviour, whereby the varying valence electron concentration (alloying) can, to a very good approximation, be modelled by simple changes to the perfect lattice calculations. This rigid-band-like behaviour is one of the reasons that the work by Jhi *et al.*¹ was possible.

What Jhi *et al.* have shown is that there is a correlation between the behaviour of the elastic constants and, by implication, the hardness, with the filling of electron-bonding states formed by the carbon and nitrogen *p*-electrons and the transition-metal *d*-electrons. In greater detail, there is competition between the filling of a set of states favourable for elastic strength and a set of states having the opposite effect. This competition results in a maximum in elastic constants such as the shear modulus — which measures the ability of a material to recover from transverse stress — as a function of electron filling, in direct correlation with a similar maximum for the hardness that has been observed in experiments. In particular, Jhi *et al.* find that the maximum in the shear modulus occurs for a valence electron concentration corresponding to a carbon or nitrogen vacancy concentration of 12%, in agreement with material testing.

Particularly interesting is that the bulk modulus of the carbonitrides, which is determined by a homogeneous volume deformation of the crystal, does not show any maximum as a function of valence electron concentration. The authors point out that as much as we like to think of the carbonitrides, with respect to their covalent-like charge density and bonding, as being analogous to covalent semiconductors, there are differences. In semiconductor or insulator alloys the hardness and the bulk and shear moduli vary in the same way with electron concentration. But the existence of two different types of energy bands for the least tightly held electrons in carbonitrides leads to different responses to shear and volume deformations.

Materials theorists are making huge progress in relating macroscopic properties of materials to the fundamental atomic constituents of matter. Both the complexity and the understanding revealed by studies such as that by Jhi *et al.*¹ illustrate that, along the road to 'materials by design', we can anticipate traversing some interesting terrain. □

Barry M. Klein is in the Department of Physics, University of California, One Shields Avenue, Davis, California 95616, USA.
e-mail: bmklein@ucdavis.edu

1. Jhi, S.-H., Ihm, J., Louie, S. G. & Cohen, M. L. *Nature* **399**, 132–134 (1999).
2. Toth, L. E. *Transition Metal Carbides and Nitrides* (Academic, New York, 1971).
3. Klein, B. M. & Papaconstantopoulos, D. A. *Phys. Rev. Lett.* **32**, 1193–1196 (1974).
4. Papaconstantopoulos, D. A., Pickett, W. E., Klein, B. M. & Boyer, L. L. *Nature* **308**, 494–495 (1984).
5. Kohn, W. & Vashishta, P. in *Theory of the Inhomogeneous Electron Gas* (eds Lundqvist, S. & Marsh, N. H.) 79–147 (Plenum, New York, 1983).
6. Singh, D. J. *Planewaves, Pseudopotentials, and the LAPW Method* (Kluwer, Boston, 1994).
7. Zunger, A. in *Statics and Dynamics of Alloy Phase Transitions* NATO ASI Series B Vol. 319 (eds Turchi, P. E. A. & Gonis, A.) 361–419 (Plenum, New York, 1994).

Photosynthesis

Mixed metabolism in plant pools

Peter D. Moore

Assemblages of plants present ecologists with some interesting conceptual problems. How can a number of species coexist when they are tapping the same set of environmental resources? All require light energy, plus a source of atmospheric (or dissolved) carbon, together with water and a fairly standard range of elements for their growth and development. So, cohabitation must depend on the plants' capacity to tap these resources in different ways, coupled with their ability to cope with the stresses that any particular habitat may present. Although such issues have been studied in terrestrial habitats, aquatic ecosystems — particularly ephemeral (or transient) aquatic habitats — have been neglected. But Jon Keeley of the United States Geological Survey Biological Resources Division in California has begun to redress the balance. Reporting in *Functional Ecology*¹, he describes his latest findings on the various photosynthetic processes found in the vegetation of temporary pools.

Keeley has looked at the photosynthetic ecology of shallow pools for many years. In 1981, he discovered a quillwort, *Isoetes howellii*, a small aquatic pteridophyte which, unexpectedly, showed the characteristics of crassulacean acid metabolism (CAM)². This photosynthetic system involves the accumulation of carbon using the enzyme phosphoenolpyruvate carboxylase as the initial stage in the fixation process, rather than the more usual ribulose 1,5-bisphosphate carboxylase-oxygenase (which results in the formation of 3-carbon products, leading to the term C₃ plants for those that use this system; Fig. 1). Crassulacean acid metabolism is most frequently associated with arid, terrestrial conditions — the short-term storage of fixed carbon in the form of organic acids allows plants to accumulate carbon during the night, enabling them to conserve water by keeping their stomata (the pores through which they take in CO₂ and, consequently, lose water) closed during the day.

The potential advantage of CAM for an

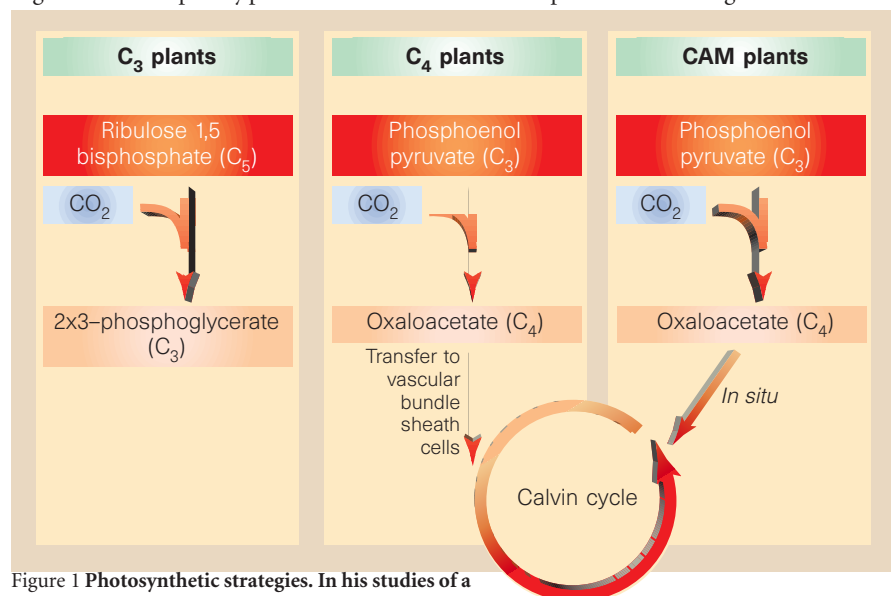


Figure 1 Photosynthetic strategies. In his studies of a transient shallow pool, Jon Keeley¹ has found that aquatic plants use a variety of photosynthetic strategies, allowing them to coexist and make best use of the available resources.

aquatic plant lies in the high affinity of phosphoenolpyruvate carboxylase for CO_2 . This allows the plant to compete for dissolved CO_2 more efficiently than C_3 plants when concentrations fall to low levels. It also permits the plant to continue fixing CO_2 during the night, when the concentration of carbon in the water rises. Low concentrations of CO_2 during the day are relatively common in shallow water — photosynthesis by macrophytes (larger aquatic plants) and algae quickly exhausts the supply of carbon, as CO_2 diffuses much more slowly through water than in the atmosphere.

Keeley has now¹ examined components of the vegetation assemblage from a pool in California. The pool has been artificially maintained over the past ten years by allowing it to fill with rainwater in January and then to dry out in mid-April (this is termed a vernal pond). Twenty plant species, which were originally introduced from a natural vernal pond, have become established under these conditions, and Keeley has examined the 15 most common of them, both structurally and for their photosynthetic properties.

Keeley found that plants with floating leaves are normally C_3 . A CAM system would be of no advantage to these plants, because their stomata are in contact with the atmosphere, yet they are not short of water. One exception is the grass *Orcuttia californica*, which has aquatic, floating and terrestrial leaves, but uses a C_4 system of photosynthesis (which also employs phosphoenolpyruvate carboxylase as the initial carbon-fixing enzyme). The submerged leaves have all the carbon-scavenging properties of the CAM system, whereas aerial C_4 leaves are particularly efficient at high temperatures and high light intensity. The two quillworts studied, *I. howellii* and *I. orcuttii*, use CAM photosynthesis, but they increase their capacity for accumulating carbon by using their roots to take up CO_2 from the sediment, where respiration by invertebrates and microbes enriches the supply. In this way, these plants tap a supply of carbon that is not immediately available to the rest of the photosynthetic foliage.

Keeley also discovered that other aquatic plant species, such as *Eleocharis acicularis*, can act as either C_3 or C_4 plants. The C_4 mechanism has the advantage over C_3 in that the plants can avoid photorespiration (which leads to inefficient carbon fixation in the presence of oxygen). If the concentration of dissolved carbon is low and that of dissolved oxygen is high, it could be valuable to avoid photorespiration. As the pool dries out, however, the diurnal depletion of the carbon supply becomes far less extreme, so the value of the C_4 mechanism declines and *E. acicularis* switches from C_4 to C_3 photosynthesis. But if conditions become very dry, especially in the heat of a California summer, the C_4

mechanism may again be useful as a means of water economy.

Several of the amphibious plant species showed marked changes in morphology, such as tougher stems and thicker cuticles, as the pool dried out and conditions changed from aquatic to terrestrial. This finding shows that these plants can both tolerate and compensate for the new stresses of a dry life. In general, the species that showed such morphological changes on drying were those that had evolved from terrestrial ancestors relatively recently. Species with a long aquatic pedigree were less morphologically plastic. So, the diversity of responses to environmental fluctuations involves both physiological and morphological modifications.

Long-term studies of this aquatic system have shown that there is a considerable inter-annual variation in the composition of the pool's plant community, probably associated with year-to-year variation in the climatic conditions. Keeley believes that this variability itself may help to promote coexistence of

the various plants present. Here is an ecosystem in which inherent instability, and a degree of unpredictability, generates diversity among its components. It is a system in which niche diversification among plants, although subtle and somewhat occult, can be discerned in the way that the available resources are apportioned³. Given the ease with which this relatively simple ecosystem can be manipulated experimentally, it is surprising that it has not previously been used as a model to study the assemblage rules of plant communities — for the variety and flexibility of photosynthetic strategies here seems to be the key to their continued existence. □

Peter D. Moore is in the Division of Life Sciences, King's College, Campden Hill Road, London W8 7AH, UK.

e-mail: peter.moore@kcl.ac.uk

1. Keeley, J. E. *Funct. Ecol.* **13**, 106–118 (1999).
2. Keeley, J. E. *Am. J. Bot.* **68**, 420–424 (1983).
3. Crawford, R. M. M. (ed.) *Plant Life in Aquatic and Amphibious Habitats* (Blackwell, Oxford, 1987).

Excitatory synapses

Neither too loud nor too quiet

Dimitri M. Kullmann

A striking feature of many excitatory synapses in the central nervous system is that the neurotransmitter glutamate simultaneously activates receptors with very different pharmacological properties. Notably, AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors open and close rapidly and mediate fast signalling, whereas NMDA (*N*-methyl-D-aspartate) receptors have slow, voltage-dependent kinetics, are permeable to calcium, and are central to the induction of long-term synaptic plasticity. On page 151 of this issue, a report by Mainen *et al.*¹ challenges a widespread assumption about how glutamate activates NMDA receptors.

For AMPA and NMDA receptors to detect the release of glutamate from presynaptic terminals, the concentration of this neurotransmitter in the synaptic cleft must be enough to activate them. Accurate estimates of the glutamate concentration after exocytosis have been obtained by measuring the sensitivity of AMPA or NMDA-mediated synaptic signals to rapidly dissociating antagonists — the higher the concentration of glutamate, the more easily the antagonist is displaced from the receptors². These (and other) experimental approaches have converged on a picture where AMPA receptors are activated by the release of a single vesicle, although the receptors may not be saturated^{3–5}. But NMDA receptors are predicted to be almost completely occupied, the difference being explained by their 100-fold higher affinity for glutamate⁶. This conclusion

agrees with the results of experiments that simulated diffusion of a few thousand glutamate molecules in the sticky and tortuous environment of the synaptic cleft⁷.

The report by Mainen *et al.*¹ comes as a surprise, because it concludes that NMDA receptors are, in fact, far from saturated by a single release event. The authors used two-photon laser scanning microscopy to measure changes in the intracellular concentration of Ca^{2+} in pyramidal neurons from the hippocampus⁸. They filled individual neurons with a high-affinity Ca^{2+} indicator, then held them at a positive membrane potential using a patch-clamp pipette. Stimulation of the presynaptic axons triggered Ca^{2+} signals that could be resolved at the level of individual dendritic spines. These Ca^{2+} signals were blocked by antagonists to NMDA receptors, but were not affected when the intracellular Ca^{2+} stores were manipulated. Importantly, in a proportion of trials the presynaptic stimuli produced no postsynaptic response (a 'failure'). This implies that the Ca^{2+} signals reflected the activation of the NMDA receptors by the probabilistic release of individual quanta of glutamate.

The authors used a simple strategy to test whether the NMDA receptors were saturated with glutamate. Given that glutamate dissociates extremely slowly⁹, the receptors should not respond to a second pulse of the neurotransmitter, delivered after a brief delay, if the first pulse is enough to saturate them. So, the Ca^{2+} signal after two stimuli

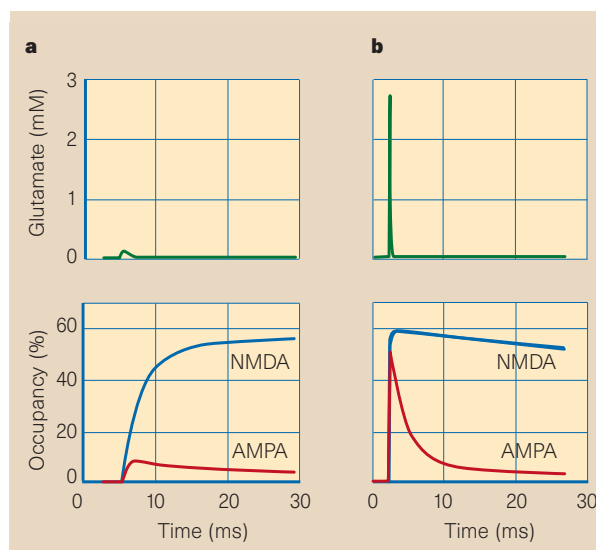


Figure 1 Occupancy of AMPA and NMDA receptors after release of the neurotransmitter glutamate. **a**, A relatively prolonged but low-concentration pulse of glutamate leads to very low occupancy of AMPA receptors (defined as the proportion of doubly bound receptors). **b**, A much briefer, high-concentration pulse gives a much higher peak occupancy of AMPA receptors, approaching that of NMDA receptors. The two different glutamate pulses give similar peak occupancies of NMDA receptors (~56%). (For details of simulations see ref. 7.)

delivered in rapid succession should be no larger than that after one pulse — assuming that either or both stimuli gave rise to glutamate release. Transmission failures could be easily identified and excluded from the analysis. Contrary to expectation, the Ca^{2+} signal was consistently larger with two stimuli than with one. So, in some trials, glutamate must have been released in response to both stimuli, allowing more of the NMDA receptors to bind glutamate than when only one stimulus was delivered. In fact, after a single release, Mainen *et al.* estimate that the receptors were less than 60% occupied.

If NMDA receptors are incompletely occupied after transmitter release, how do AMPA receptors respond at all, given that their affinity for glutamate is so much lower? Should we expect to see synaptic events mediated by NMDA but not AMPA receptors? Synaptic signals mediated exclusively by NMDA receptors are, in fact, readily seen at the sub-physiological temperatures under which most *in vitro* experiments are done. And the finding that AMPA receptors often fail to detect transmitter release under these conditions has prompted the term ‘silent synapses’. But at near-physiological temperatures — the conditions monitored by Mainen *et al.* — such silent synapses are rarely observed unless glutamate transporters are blocked¹⁰. This result implies that glutamate uptake has a critical temperature-dependent role in determining receptor occupancy³.

How might the speed at which glutamate is cleared explain how both NMDA and AMPA receptors contribute to synaptic signalling at physiological temperatures? The answer probably lies with the detailed kinetic properties of the two types of receptor. The NMDA receptors are less sensitive to glutamate when it is present for only a very short time than when it is present for a longer time. Indeed, by simulating the response of the two receptors to a very brief pulse — as may occur after exocytosis *in vivo* — the peak

occupancies of both receptors turn out to be very similar (Fig. 1). This means that both NMDA and AMPA receptors may detect transmitter release to roughly the same extent. At sub-physiological temperatures, on the other hand, glutamate persists for longer owing, in part, to slower uptake, and NMDA receptors become relatively more sensitive to transmitter release than AMPA receptors. To test this hypothesis, we need more studies into the release, diffusion and uptake of glutamate in the extracellular space, and a detailed comparison of glutamate-receptor kinetics at different temperatures.

Whatever the outcome of the debate over the relative occupancy of AMPA and NMDA receptors, the striking conclusion of Mainen and colleagues’ work is that a single release event falls roughly in the middle of the dynamic (response) range for a cluster of NMDA receptors. The lack of saturation is surprising, because it should make synapses less reliable than previously thought. On the other hand, the phenomenon illustrates a ubiquitous biological principle — that input is matched to output. This arrangement maximizes the sensitivity of the postsynaptic neuron to changes in the amount of glutamate released. □

Dimitri M. Kullmann is in the Department of Clinical Neurology, Institute of Neurology, Queen Square, University College London, London WC1N 3BG, UK.

e-mail: dkullman@ion.ucl.ac.uk

1. Mainen, Z. F., Malinow, R. & Svoboda, K. *Nature* **399**, 151–155 (1999).
2. Clements, J. D. *Trends Neurosci.* **19**, 163–171 (1996).
3. Tong, G. & Jahr, C. E. *Neuron* **13**, 1195–1203 (1994).
4. Silver, R. A., Cull-Candy, S. G. & Takahashi, T. J. *Physiol. (Lond.)* **494**, 231–250 (1996).
5. Liu, G., Choi, S. & Tsien, R. W. *Neuron* **22**, 395–409 (1999).
6. Patneau, D. K. & Mayer, M. L. *J. Neurosci.* **10**, 2385–2399 (1990).
7. Rusakov, D. A. & Kullmann, D. M. *J. Neurosci.* **18**, 3158–3170 (1998).
8. Svoboda, K. & Mainen, Z. F. *Neuron* **22**, 427–430 (1999).
9. Lester, R. A. *et al.* *Nature* **346**, 565–567 (1990).
10. Asztely, F., Erdemli, G. & Kullmann, D. M. *Neuron* **18**, 281–293 (1997).

Daedalus

Cultured diamonds

One of the best ways of purifying a chemical substance is by crystallization. A growing crystal accepts molecules that fit its lattice, and rejects those that do not, so a pure product will crystallize from even quite an impure solution. Indeed, even thermodynamically unstable crystals will grow from a solution, if a few ‘seed crystals’ are added to give them a start.

The most valuable of all thermodynamically unstable crystals is, of course, diamond. So Daedalus wants to use the crystallization principle to grow big, pure diamonds. His first idea was to dissolve buckminsterfullerene (the only soluble form of carbon) in benzene or liquid xenon, add a few seed crystals of diamond, and let the solution evaporate. But it seems unlikely that a complete C_{60} molecule would rearrange to a diamond lattice merely by colliding with a diamond surface. A smaller carbon unit is required.

The smallest stable carbon unit is the C_2 radical. It is most easily generated by heating carbon in a vacuum. The hot surface slowly evaporates by emission of these radicals — which is why the old carbon-filament lamps had such a limited life. A C_2 radical hitting a diamond surface should combine with it to extend the diamond lattice. So a steady rain of such radicals should, by the crystallization principle, grow a diamond of any size from a small initial seed.

Daedalus’s diamond-growing apparatus has a source of C_2 radicals (a carbon filament or arc) surrounded by a heated precious-metal nozzle to direct them as a beam towards an array of seed diamonds on a cooled platform. The whole thing is in a vacuum chamber. The seed diamonds are first cleaned by argon-ion bombardment, to remove surface contamination and expose their pristine lattice surfaces; the C_2 beam is then turned on to grow them to the required size. By narrowing and steering the beam, and orienting the diamonds under it, they can be shaped as well.

Cheap diamond-growing will transform the industry. Diamond jewellery will lose chic and value, while industrial and scientific diamonds will proliferate in the form of cutting tools, optical windows, heat-sinks and so on. But Daedalus muses that nature may have got there first. Many ‘carbon stars’ shower carbon vapour out into the interstellar vacuum. If a few seed diamonds happen to be orbiting in the vicinity, they might over millions of years grow into diamond planets.

David Jones

Myopia and ambient lighting at night

Myopia, or short-sightedness, occurs when the image of distant objects, focused by the cornea and lens, falls in front of the retina. It commonly arises from excessive postnatal eye growth, particularly in the vitreous cavity. Its prevalence is increasing and now reaches 70–90% in some Asian populations^{1,2}. As well as requiring optical correction, myopia is a leading risk factor for acquired blindness in adults because it predisposes individuals to retinal detachment, retinal degeneration and glaucoma. It typically develops in the early school years but can manifest into early adulthood². Its aetiology is poorly understood but may involve genetic and environmental factors^{1,2}, such as viewing close objects, although how this stimulates eye growth is not known³. We have looked at the effects of light exposure on vision, and find a strong association between myopia and night-time ambient light exposure during sleep in children before they reach two years of age.

Research in species as diverse as chicks and monkeys indicates that postnatal eye growth and refractive development are governed by a vision-dependent retinal mechanism acting mainly within the eye, with only limited participation of the brain and extra-ocular neural pathways^{3,4}. The duration of the daily light period has been shown to affect eye growth in chicks⁵, so we investigated whether refractive development in children might associate with any

recognizable pattern of light exposure. Because early neonatal visual experience markedly affects refractive development in animals^{4–6}, we evaluated light exposure both at the child's present age and before the age of two years, a period during which the eye grows rapidly⁷ but before the usual onset of myopia².

Between January and June 1998, parents of children aged 2–16 years (median age 8.0 years; $n = 479$ children, 55% males; 70% Caucasian, 30% African-American, less than 1% Asian-American) that were seen as outpatients in a university paediatric ophthalmology clinic completed a questionnaire on the child's light exposure both at present and before the age of two years. Children with amblyopia, cataract, glaucoma or a history of prematurity were excluded.

The prevalence of myopia and high myopia during childhood was strongly associated with ambient light exposure during sleep at night in the first two years after birth (Fig. 1). The relation between refraction and night-time light was dose dependent, as a greater proportion of children became myopic if they slept at night during their first two years with room lighting rather than with a night light. The increased prevalence of myopia resulted from a smaller proportion of emmetropic children, as light exposure did not relate to the proportion of hyperopic children. We found no other association of refraction with report-

ed light exposure, including no relation with night-time lighting at the child's present age.

An influence of ambient lighting during sleep on refractive development is plausible, because eyelids of human adults and infants transmit some visible light, mostly at longer wavelengths⁸. The scotopic retinal sensitivity of infants is relatively good compared with that of adults, particularly by the age of 18 weeks⁹. Further, sutured eyelids of infant monkeys transmit a degraded image and perturb refractive development⁶.

This study does not indicate whether early visual experience influences ocular anatomy by age two or only later, and does not permit conclusions to be made about the timing of the onset or progression of myopia. It raises the possibility of a 'critical period' for refractive development analogous to that for visual function¹⁰.

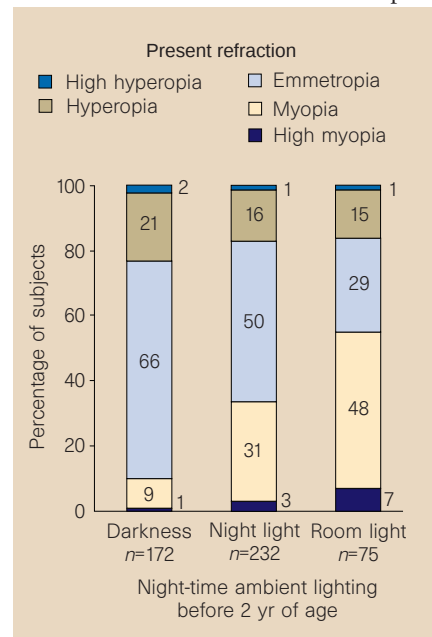
Although it does not establish a causal link, the statistical strength of the association of night-time light exposure and childhood myopia does suggest that the absence of a daily period of darkness during early childhood is a potential precipitating factor in the development of myopia. The results are further qualified by the limitations of collecting behavioural data by questionnaire and the lack of information on whether parental lighting preferences independently associate with other factors known to be correlated with myopia, such as parental socio-economic or refraction status^{1,2}. The generalizability of this relationship, observed in a tertiary referral centre, also requires extension to other populations and especially to Asian groups, which are severely affected by myopia¹. Despite these qualifications, it seems prudent that infants and young children sleep at night without artificial lighting in the bedroom, while the present findings are evaluated more comprehensively.

**Graham E. Quinn*†, Chai H. Shin*,
Maureen G. Maguire†, Richard A. Stone†**

*Department of Ophthalmology,
Children's Hospital of Philadelphia,
University of Pennsylvania School of Medicine,
Philadelphia, Pennsylvania 19104, USA

†Department of Ophthalmology,
Scheie Eye Institute,
University of Pennsylvania School of Medicine,
Philadelphia, Pennsylvania 19104, USA
e-mail: stone2@mail.med.upenn.edu

Figure 1 Present refractions of children of ages 2–16 yr and night-time light exposure before the age of 2 yr. The prevalence of myopia increases markedly with increased levels of night-time ambient lighting during sleep before the age of 2 yr. On a questionnaire approved by an institutional review board, parents were asked, "Under which lighting condition did/does your child sleep at night?" before the age of 2 yr and at present; they chose between 'room lighting', 'a night light' (typically, in the USA, a dim socket-mounted fixture of ~4 W) and 'darkness'. Other questions addressed the lighting in various rooms at home, lighting at day care or school, geographical locations where the child had lived and current use of sunglasses. On the basis of the mean cycloplegic spherical equivalent of both eyes at the child's most recent ophthalmic examination, we separated the refractions into five groups: high hyperopia (long-sightedness), $\geq +5.0$ dioptres (D); hyperopia, $+2$ to $< +5$ D; emmetropia ('normal' childhood refraction), $< +2$ to < -0.5 D; myopia, -0.5 to < -5.0 D; high myopia, ≤ -5.0 D. The percentage of children in the combined myopia and high myopia groups at their present age increased with increasing night-time light exposure before the age of 2 yr (χ^2 with 1 degree of freedom = 55.1, $P < 0.00001$). The strength of the relation was maintained after adjustment for age by logistic regression analysis. The same relation held for separate analyses of the Caucasian and African-American subjects ($P < 0.00001$ for each group; results not shown).



1. Saw, S.-M., Katz, J., Schein, O. D., Chew, S.-J. & Chan, T.-K. *Epidemiol. Rev.* **18**, 175–187 (1996).
2. Zadnik, K. & Mutti, D.O. in *Borish's Clinical Refraction* (ed. Benjamin, W. J.) 30–46 (Saunders, Philadelphia, 1998).
3. Stone, R.A. in *Myopia Updates: Proceedings of the 6th International Conference on Myopia* (ed. Tokoro, T.) 241–254 (Springer, Tokyo, 1997).
4. Wallman, J. *Prog. Retinal Res.* **12**, 133–153 (1993).

5. Stone, R. A., Lin, T., Desai, D. & Capehart, C. *Vision Res.* **35**, 1195–1202 (1995).
6. Raviola, E. & Wiesel, T. N. *N. Engl. J. Med.* **312**, 1609–1615 (1985).
7. Larsen, J. S. *Acta Ophthalmol. (Copenh.)* **49**, 873–886 (1971).
8. Robinson, J., Bayliss, S. C. & Fielder, A. R. *Vision Res.* **31**, 1837–1840 (1991).
9. Hansen, R. M. & Fulton, A. B. in *Early Visual Development, Normal and Abnormal* (ed. Simons, K.) 130–142 (Oxford Univ. Press, New York, 1993).
10. Daw, N. J. *Arch. Ophthalmol.* **116**, 502–505 (1998).

Superconductors under stress

Locquet *et al.*¹ reported a doubling of the critical temperature (T_c) of the superconductor $\text{La}_{1-x}\text{Sr}_x\text{CuO}_4$, from 25 to 49 K, in compressively strained thin films on a SrLaAlO_4 substrate. This led to speculation that values of T_c close to or even beyond 200 K could be obtained without excessive external pressure, starting from an unstrained cuprate with high T_c (> 100 K). The stress would be applied to the ab (CuO_2) plane of this system of tetragonal symmetry. Such expectations are theoretically unrealistic.

To understand the phenomenon they observed, Locquet *et al.* cited data on uniaxial pressure gradients of T_c on the same compound^{2,3}. In particular, they inferred from the negative sign of the quantity $dT_c/d\epsilon_c$, where ϵ_c is the strain (minus the relative change in distances along the c -axis), that elongation (negative strain) of the c -axis increases the critical temperature, if atomic coordinates in the ab plane are kept constant.

Without questioning the validity of their experimental results, we disagree with Locquet *et al.* on two points. First, the higher the critical temperature in the unstrained state, the closer is the system with respect to optimum density in coordinate space, or to the optimum level of the Fermi energy. Quenching will then lead to a much weaker effect than at low T_c , and may even lower the critical temperature. This conclusion reflects *a priori* on results selecting the Hg-cuprate with the present record T_c of 134 K (three CuO_2 layers per molecular unit).

Second, the cuprates $\text{La}_{2-x}\text{M}_x\text{CuO}_4$ (where M is Sr or Ba) are unsuited as a reliable source of microscopic information via stress-strain relations because of observed structural distortions, phase transitions, non-rigid tilting of CuO_6 octahedra, and so on³, under external parameters and doping. As a result, conventional stress-strain relations, which are macroscopic and phenomenological, are not applicable to these systems on an atomic scale. This implies that the usual equalities (all derivatives are partial), $dT_c/d\epsilon_i = \sum_j (dT_c/dP_j)C_{ji}$, where ϵ_i is

the strain along the i th axis and C_{ji} are elastic moduli, cannot provide the information desired. With i denoting the c -axis direction, the right-hand side is large and negative^{3,4} for $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$, but its indicated relation to a quantity which presupposes that all coordinates except those along the c -axis remain unchanged does not hold.

Locquet *et al.* incorrectly conclude that these negative values imply an increase of T_c brought about solely by enlarging the c -axis. In particular, their criticism of the interlayer tunnelling model proposed by Anderson *et al.*⁴ is irrelevant, although this model has been criticized on other grounds⁵.

Laurens Jansen*, Ruud Block†

*c/o Institute of Theoretical Physics,
ETH-Hönggerberg HPZ 8093,
Zürich, Switzerland

e-mail: jansen@itp.phys.ethz.ch

†J. H. van't Hoff Institute,
University of Amsterdam,
1018 WV Amsterdam, The Netherlands

1. Locquet, J.-P. *et al. Nature* **394**, 453–456 (1998).
2. Gugenberger, F. *et al. Phys. Rev. B* **49**, 13137–13142 (1994).
3. Nohara, M. *et al. Phys. Rev. B* **52**, 570–580 (1995).
4. Chakravarty, S., Sudbo, A., Anderson, P. W. & Strong, S. *Science* **261**, 337–340 (1993).
5. Tsvetkov, A. A. *et al. Nature* **395**, 360–362 (1998).

Polar gigantism dictated by oxygen availability

The tendency of some animals to be larger at higher latitudes ('polar gigantism') has not been explained, although it has often been attributed to low temperature and metabolism¹. Investigation of gigantism requires widely distributed taxa with extensive species representation at many well-studied sites. We have analysed length data for 1,853 species of benthic amphipod crustaceans from 12 sites worldwide, from polar to tropical and marine (continental shelf) to freshwater environments. We find that maximum potential size (MPS) is limited by oxygen availability.

Size spectra are right-skewed at all sites, but skewedness increases as temperature decreases (Fig. 1). Analysing gigantism needs emphasis on right-hand extremes of distributions. However, potential sampling bias at some sites precludes using absolute maximum size, so we used the threshold size separating the smallest 95% of species from the largest 5% (which we refer to as $\text{TS}_{95/5}$). For marine sites, $\text{TS}_{95/5}$ increases as water temperature decreases. Sites with low salinity, such as Lake Baikal (0 practical salinity units, p.s.u.), Caspian (13 p.s.u.) and the Black Sea (17 p.s.u.) lie above the marine relationship, and the discrepancy increases as salinity decreases (Fig. 2a), with the Lake Baikal $\text{TS}_{95/5}$, for example, being 1.8 times the marine value at the same tem-

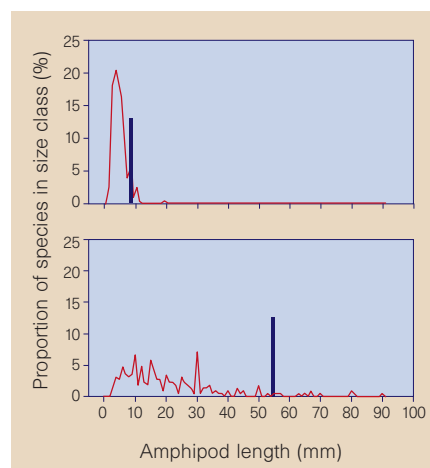


Figure 1 Amphipod size spectra for the two sites (Madagascar³, top; Lake Baikal², bottom) for which the maximum and minimum $\text{TS}_{95/5}$ values, indicated by bars, were obtained. In total, data were obtained from 12 sites worldwide: Madagascar, mean annual water temperature (T) = 25 °C, $\text{TS}_{95/5}$ = 8.3, number of species at each site (n) = 314; Mediterranean Sea, T = 19 °C, $\text{TS}_{95/5}$ = 12.6, n = 350; Black Sea, T = 16.5 °C, $\text{TS}_{95/5}$ = 20.3, n = 93; Caspian Sea, T = 15 °C, $\text{TS}_{95/5}$ = 25.4, n = 69; British Isles, T = 11 °C, $\text{TS}_{95/5}$ = 23.2, n = 172; Magellanic region, T = 9.75 °C, $\text{TS}_{95/5}$ = 23.4, n = 164; subantarctic islands, T = 9.5 °C, $\text{TS}_{95/5}$ = 21.9, n = 181; Lake Baikal, T = 6 °C, $\text{TS}_{95/5}$ = 54.4, n = 226; Barents Sea, T = 4 °C, $\text{TS}_{95/5}$ = 35.4, n = 134; South Georgia, T = 1.5 °C, $\text{TS}_{95/5}$ = 40.5, n = 150; West Antarctica, T = 0.75 °C, $\text{TS}_{95/5}$ = 41.5, n = 297; East Antarctica, T = 0.0 °C, $\text{TS}_{95/5}$ = 43.6, n = 195). Analyses were restricted to a depth of 250 m (continental shelf depth), except for Antarctica, which included species to 500 m because of continental shelf depression by the Antarctic icecap. Sites that had fewer than 50 species described were not analysed.

perature. But oxygen solubility increases as salinity decreases, and replotting $\text{TS}_{95/5}$ against water-dissolved oxygen content removes this discrepancy and produces a linear relation: $\text{TS}_{95/5} = -42.6 + 0.252 \text{ O}_2$ (n = 12; r^2 = 0.98; F = 51.69; P < 0.0001) (Fig. 2b).

Thus, oxygen availability controls $\text{TS}_{95/5}$. The fit of regressions (not shown) relating lower threshold size values (such as $\text{TS}_{90/10}$ and $\text{TS}_{50/50}$) with oxygen improves with increasing threshold size value, indicating that oxygen becomes more important relative to other ecological factors as size increases. Furthermore, the relation between $\text{TS}_{95/5}$ and temperature for marine sites is curvilinear (Fig. 2a), reflecting the nonlinear relation between seawater oxygen content and temperature.

$\text{TS}_{95/5}$ reaches zero when mean environmental oxygen is 183 μmol per kg of water. This should indicate an environmental limit for amphipods. A variety of hot and/or highly saline conditions would produce 183 μmol O_2 per kg, but these areas are inhabited by brine shrimp and ostrac-

cods, and not by amphipods (S. Ruffo, personal communication).

Minimum size does not vary significantly with environmental oxygen (Pearson correlation coefficient = 0.402, $P = 0.195$). Thus, MPS increases dramatically with oxygen, modal size increases less, and minimum size does not increase at all. The overall effect widens the size spectrum, indicating that oxygen availability is not the same selective pressure for all species, but rather sets upper limits to MPS.

In ectotherms, metabolic rate increases with temperature, increasing tissue maintenance costs². Although MPS decreases with increased temperature for marine sites, the largest amphipods were not found in the coldest sites (high Antarctic, 0 °C), but in Lake Baikal (+6 °C). Temperature-dependent tissue synthesis and catabolism trade-offs do not limit MPS, as the observed increases in MPS at freshwater sites, despite greater osmoregulatory costs, would require unfeasibly large increases in resource acquisition.

Oxygen enters amphipod blood through a low-efficiency gill³ and is transported both as dissolved oxygen and bound to haemocyanins. Marine amphipod haemolymph contains 10–20 mg ml⁻¹ haemo-

cyanin⁴, which is low for crustaceans. Polar amphipods are thought to be similar to the Antarctic giant isopod *Glyptonotus antarcticus*, which also has relatively little haemocyanin⁵ and carries 60–70% of its circulating oxygen in solution. The amount of oxygen carried in solution is thought to be 30–40% in tropical amphipods and 60–70% in Lake Baikal species, and differs because marine amphipod haemolymph is isotonic with seawater (950–1,150 mosM), whereas freshwater species are hypertonic (340–360 mosM) (I. Zerbst-Boroffka, personal communication). Amphipod haemocyanins may also be more important for osmoregulation than oxygen transport⁶.

The amount of oxygen dissolved in saturated haemolymph in amphipods from Lake Baikal at 6 °C is similar to that of marine species at 0 °C, and species from Lake Baikal and the Antarctic should be similar sizes. $TS_{95/5}$ is 54.4 mm for Baikal amphipods and 43.6 mm for Antarctic species. This suggests that saturation levels of dissolved oxygen in the haemolymph increase from tropical to polar environments, and from marine to freshwater environments, with gill efficiency being the critical factor determining MPS.

The difference in partial pressure between the external medium and circulating haemolymph determines oxygen uptake across the gills, according to Fick's law. For similar external partial pressure and efficiency, more oxygen enters the blood at sites with low temperature and salinity because the absolute oxygen concentration is higher. The saturation level and absolute oxygen concentration in the blood are also higher, allowing increased size because a greater oxygen mass will increase the possible path length of the circulatory system.

We obtained strong relations between external oxygen concentration and length, not body mass. However, amphipods, like several groups exhibiting low-temperature gigantism (such as pycnogonids and nemerteans), have restricted circulatory systems with few lateral branches. In other groups, stronger relations with body mass would be expected. Whether the MPS is reached will depend on several factors. In environments with many species, selection pressures will drive niche exploitation and a range of sizes will be represented.

Oxygen supply may also have led to insect gigantism in the Carboniferous period, because atmospheric oxygen was 30–35% (ref. 7). The demise of these insects when oxygen content fell indicates that large species may be susceptible to such change. Giant amphipods may therefore be among the first species to disappear if global temperatures are increased or global oxygen levels decline. Being close to the critical MPS limit may be seen as a specialization

that makes giant species more prone to extinction over geological time.

Gauthier Chapelle*, Lloyd S. Peck†

*Institut Royal des Sciences Naturelles de Bruxelles, Rue Vautier 29, B-1000 Bruxelles, Belgium

†British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK
e-mail: l.peck@bas.ac.uk

1. Atkinson, D. & Sibly, R. M. *Trends Ecol. Evol.* **12**, 235–239 (1997).
2. Ivleva, I. V. *Int. Rev. Ges. Hydrobiol.* **65**, 1–47 (1980).
3. Wolvsekamp, H. P. & Waterman, T. H. in *The Physiology of Crustacea* Vol. 1 (ed. Waterman, T. H.) 35–100 (Academic, New York, 1960).
4. Spicer, J. I. *Mar. Behav. Physiol.* **24**, 123–136 (1993).
5. Whiteley, N. M., Taylor, E. W., Clarke, A. & El Haj, A. J. *Polar Biol.* **18**, 10–15 (1997).
6. Hagerman, L., Sandberg, E. & Vismann, B. *Mar. Biol.* **130**, 209–212 (1997).
7. Graham, J. B., Dudley, R., Aguilar, N. M. & Gans, C. *Nature* **375**, 117–120 (1995).
8. Ledoyer, M. *Faune de Madagascar. Crustacés Amphipodes Gammariens* (CNRS, Paris, 1982–1986).
9. Bazikalova, A. Y. *Lake Baikal Amphipods. Proc. Baikal. Limnol. Stn Acad. Sci. USSR* **11**, 1–440 (1945).

No role for colour in symmetry perception

Bilateral colour symmetry, such as that evident in a Siberian tiger's face (Fig. 1a), is relevant to many animals^{1,2}, including humans^{3,4}. We examined the role of colour in symmetry perception by asking observers to detect colour symmetry in regular grids of coloured squares (a colour-symmetrical image has regions of the same colour located equidistantly from a vertical axis). Our results suggest, unexpectedly, that the mechanisms of symmetry perception are inherently colour-blind: although observers can verify colour symmetry, they do so only by shifting attention from one colour to the next and assessing the symmetry of regions of that colour.

Observers were shown displays that either exhibited complete colour symmetry about the vertical midline (Fig. 1b), or contained one pair of squares that were mismatched in colour. Subjects pressed a button to indicate whether the pattern was colour-symmetrical, and response times and errors were assessed. The displays were constructed with either two colours (crimson and scarlet) or four colours (green, yellow, blue and red). Colours were chosen such that every pair of colours in the four-colour displays was more easily discriminated than the pair used in the two-colour displays.

The most obvious hypothesis about colour symmetry perception, called colour matching, involves the colours of corresponding points or regions on different sides of an axis being compared at the same time. Colour symmetry is registered when

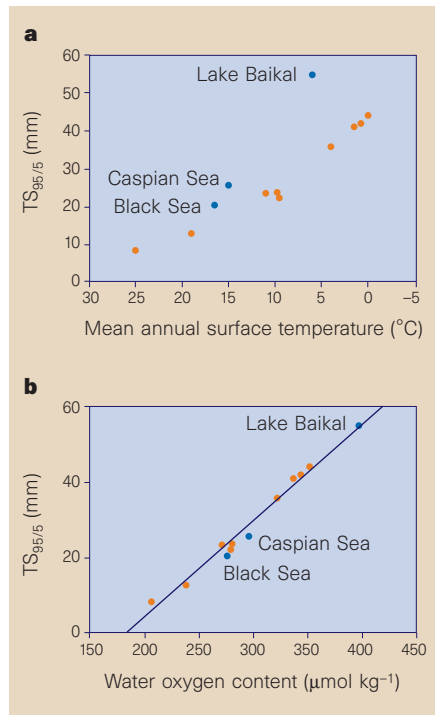


Figure 2 Effects of temperature and oxygen availability on amphipod MPS. Data are for nine marine (orange circles) and three reduced-salinity sites (blue circles). **a**, $TS_{95/5}$ plotted against mean annual water temperature (inverted scale). **b**, $TS_{95/5}$ plotted against calculated dissolved oxygen content at saturation ($\mu\text{mol per kg}$), based on surface water mean temperature and salinity. Not every habitat will experience permanent high oxygen saturation, but the 100% value represents optimal conditions for attaining large size.

every matching process is successful. This account predicts that responses to two-colour displays should be slower and more prone to errors than responses to four-colour displays, because the greater similarity of mismatched mates in the two-colour condition should slow the matching. The number of colours should in itself have no effect. But if colour symmetry is verified by attending to individual colours one at a time (the attention-switching hypothesis), the opposite should be true: response times should increase markedly with the number of colours, because each additional colour requires an additional symmetry-checking process.

The 21 subjects each began with 60 practice trials, followed by six blocks of 40 trials. A fixation cross in the screen centre preceded the display and remained until a key was depressed. Subjects were told to respond quickly and accurately. A beep sounded whenever an error was made. Displays subtended 9° of visual angle and were composed of 32 evenly spaced squares (16 elements per side), in either two or four colours. Colours were randomly assigned to squares, with the constraint that each colour was assigned to equal numbers of elements in the symmetrical displays; the colour of one square was altered to make asymmetrical displays.

Responses to four-colour displays were slower and less accurate, and responses to symmetrical displays were slower than responses to asymmetrical ones (Fig. 1c). The results support the attention-switching hypothesis and reject the more obvious colour-matching account. The mean response time to two-colour displays was 1,187 ms and that to four-colour displays was 1,961 ms; this difference was significant ($F(1,20) = 113.6$; mean squared error (MSe) = 109,585; $P < 0.0001$). The mean response time to asymmetrical displays was 1,395 ms and to symmetrical displays 1,755 ms; this difference was significant ($F(1,20) = 40.7$; MSe = 63,664; $P < 0.0001$). The overall ranking of mean response times by conditions seen in Fig. 1c was shown by 17 out of 21 observers. The mean error rate to two-colour displays was 5%, to four-colour displays 9%, to symmetrical displays 3%, and to asymmetrical displays 11%. The lower error rate for symmetrical displays suggests that the judgement 'symmetrical' is made after an exhaustive search for mismatches has failed, because mismatches are more often missed than falsely registered.

The attention-switching hypothesis makes a further, distinctive prediction for displays in which two squares differ in colour from their mates on the opposite side of the display. We used four-colour symmetrical displays, as in the previous experiment. In this experiment, however,

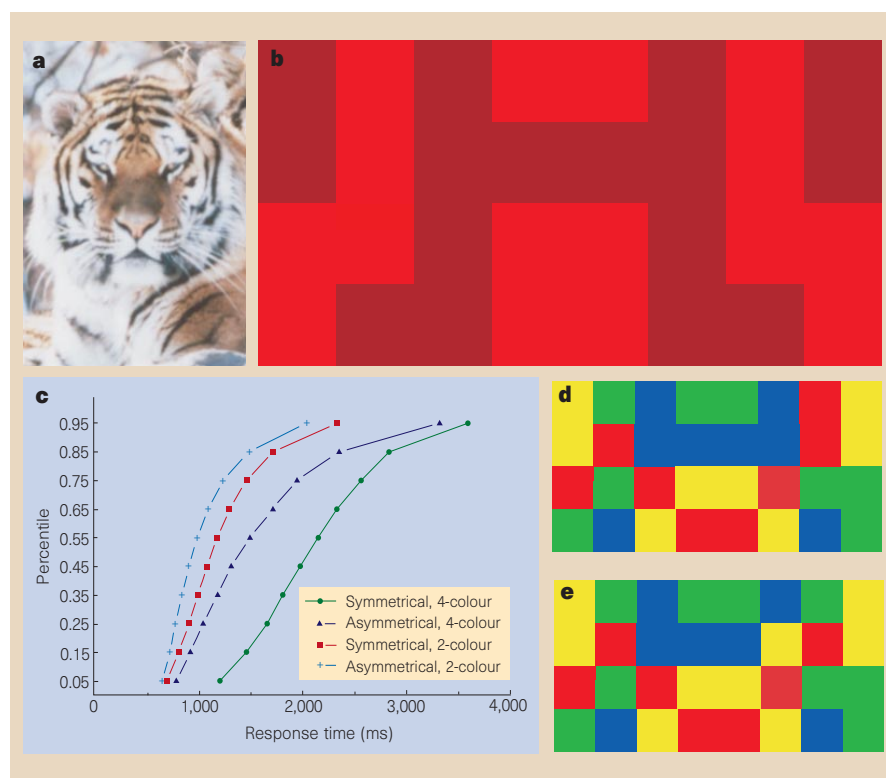


Figure 1 Colour symmetry. **a**, Siberian tiger (photographed by M. Fontaine); **b**, symmetrical two-colour display; **c**, 'Vincentized' data distributions to two-colour and four-colour symmetrical and asymmetrical displays; **d**, asymmetrical 'ABBA' two-mismatch display; **e**, asymmetrical 'ABCD' two-mismatch display.

two types of asymmetrical displays were compared, both of which included two mismatching pairs. In 'ABBA-asymmetrical' displays (Fig. 1d), both mismatched squares involved two colours, for example red and green. In contrast, in the 'ABCD-asymmetrical' displays shown in Fig. 1e, the two mismatches affected all four colours. If observers check colours sequentially, then in the ABCD display the first colour checked will always reveal an asymmetry. Thus, response times should, on average, be faster in this condition. When only two colours are involved, observers might need to check two or even three colours before encountering an asymmetry. Thus, attention switching predicts faster responses to ABCD-asymmetrical displays, whereas colour matching makes no such prediction.

Responses to ABBA-asymmetrical displays (1,449 ms) were significantly slower than to ABCD-asymmetrical displays (1,304 ms) ($F(1,15) = 23.9$; MSe = 7,071; $P < 0.0001$), supporting the attention-switching hypothesis. The ranking of mean response times (in ascending duration: ABCD-asymmetrical, ABBA-asymmetrical, and symmetrical) was shown by 13 out of 15 subjects. The error rate for ABCD-asymmetrical displays (7%) was also lower than for the ABBA-asymmetrical displays (12%) ($F(1,15) = 12.8$; MSe = 0.0016; $P < 0.003$).

The results described here suggest that, despite the apparent conspicuousness of

colour symmetry, and contrary to previous suggestions^{5,6}, symmetry detection in the human visual system is insensitive to colour in itself. When people seek to judge colour symmetry, they can do so accurately, but only by relying on a strategy of attention switching based on colour. Presumably this is done by selectively attending to one colour at a time. Evidently, then, attentional gating must arise earlier in the visual pathway than symmetry detection. This conclusion closely parallels the conclusions of Lu and Sperling about motion perception⁷. These authors reported that voluntary attention to a feature permits information to be gated before its analysis by motion-perception mechanisms. Both phenomena reinforce the emerging consensus that visual selective attention gates the flow of information early in the visual system.

Dawn Morales, Harold Pashler

Department of Psychology, 0109, University of California, San Diego, 9500 Gilman Drive, San Diego, California 92093, USA
e-mail: hpashler@ucsd.edu

- Giurfa, M., Eichmann, B. & Menzel, R. *Nature* **382**, 458–461 (1996).
- Ridley, M. *Science* **257**, 327–328 (1992).
- Mach, E. *The Analysis of Sensations and the Relation of the Physical to the Psychological* (Dover, New York, 1959). (Originally published in 1886.)
- Julesz, B. *Foundations of Cyclopean Perception* (Chicago Univ. Press, 1971).
- Troscianko, T. *Vision Res.* **2**, 547–554 (1987).
- Troscianko, T. *Perception* **14**, A24–A25 (1986).
- Lu, Z.-L. & Sperling, G. *Nature* **377**, 237–239 (1995).
- Ratcliff, R. *Psychol. Bull.* **86**, 446–461 (1979).

A fantasia of biological feminism

Woman: An Intimate Geography

by Natalie Angier

Houghton Mifflin/Little, Brown: 1999.

480 pp. \$25/£17.99

Just Like a Woman: How Gender Science is Redefining What Makes Us Female

by Dianne Hales

Virago: 416 pp. £16.99

Olivia Judson

Once upon a time, poets and musicians had a monopoly on illuminating the mystery that is woman. Then science started to creep in on the act. As the fervour for eugenics soured into revulsion, however, so too did enthusiasm for biological explanations of differences between peoples. For decades, feminists fought the notion that observed differences between men and women could be explained by biology rather than by inequality of opportunity, education and environment. They were right to do so. Stereotype after stereotype — from marathon running to performing surgery — has been dismantled.

But now biology is back. And this time feminists are interested. Unfortunately, on reading the mess they make of it, it is tempting just to hand everything back to the poets.

You see, vaginal discharge “is the lubricant beneath the illusion of carapace, reminding us that physiologically we are all aquatic organisms”, or at least so says Natalie Angier, a science journalist for *The New York Times*, in *Woman: An Intimate Geography*. She tells us that “the body is a creature of habit, and the longer it has been exposed to the chemistry of bondage, the more prone it becomes to emotional flashbacks”. She reveals that “love can feel aggressive to the point of violence. We commit our most heinous acts of aggression in the name of love. The love of God drives crusades and jihads; the love of tribe drives genocide.” But that’s OK. This terrible love, she explains, is in our nature: “we love, at bottom, because we must, for we are a sexually reproducing species.” Never mind that many sexually reproducing animals never actually meet their partners, or that plants and fungi are not known for passion, sexual liaisons notwithstanding.

Dianne Hales, also a science journalist, espouses equally silly views in *Just Like a Woman: How Gender Science is Redefining What Makes Us Female*. She reveals that “if a man, echoing Descartes, could declare, ‘I think; therefore I am’, women throughout time could have said, ‘we feel; therefore we are.’” She suggests that “so deep is the urge to reach out to others, so great is its importance to the survival of mother and child, that it

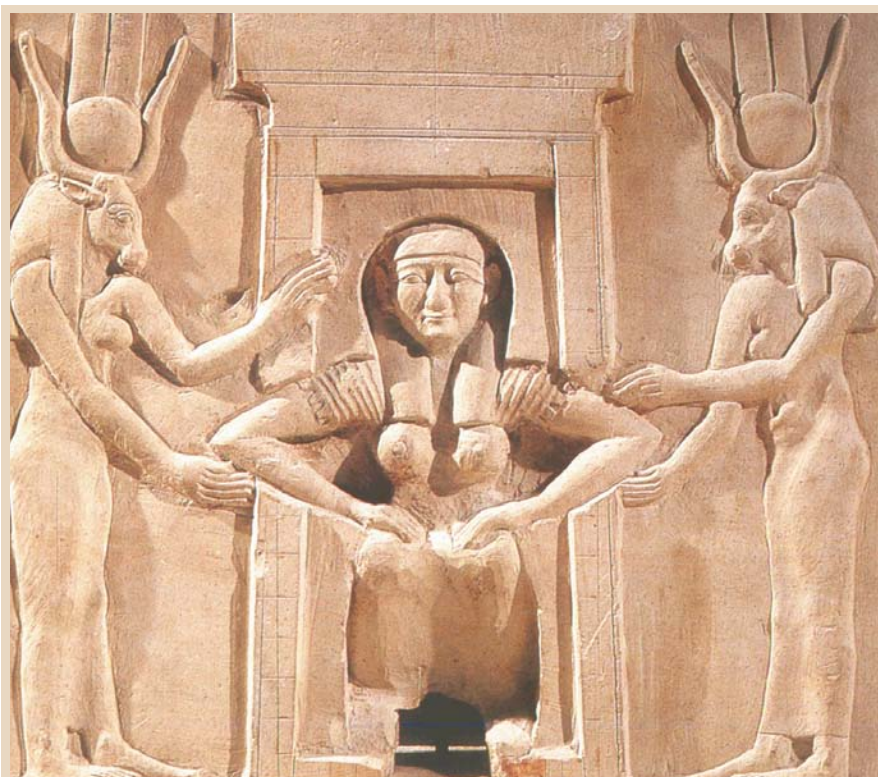
may have given rise to our seemingly instinctive need to connect”. Perhaps this is why she has come to the sinister conclusion that “women have always been more open to visitations by otherworldly beings”.

In their separate quests to fathom what makes a woman, each writer enlists grandmothers and mitochondria, eggs and oestrogen, clitorises and climaxes, breasts and blood, wombs and waists, conception and abortion, motherhood and menopause. Both would like to see a girl’s first menstrual period be a public celebration; Angier even advocates “a woman-centred myth of menstruation, a construct of our shared feminine low-giene — something on a par with the male pissing ritual, perhaps.” And naturally, both are fascinated by sex. But while Hales argues that women mostly like to be hugged and that orgasm is secondary, Angier recommends marijuana to cure anorgasmia and revels in the thought that, unlike members of the weaker sex, a sexually athletic woman can have 50 orgasms in an hour.

Hales’s book is stolid and earnest, stuffed with countless interviews and scientific results. She takes on worthy subjects, decry-

ing the inattention women have historically received in medical research, and speculating on why women are so much more likely to be depressed than men. But the book has three grave flaws. In ascending order of gravity: it is full of platitudes; the results are jumbled, with good science thrown after bad; and, finally, it is mired in evolutionary psychobabble of the most incoherent kind. For example, Hales proposes that elephant seals can show how men and women will compete differently in tough office environments. And she tells us — apparently without irony — that “females of other species have confronted many of the hassles today’s women face, including sexual harassment and unwanted pregnancy. It’s not surprising. They’ve been around a lot longer.”

Angier, meanwhile, claims her book is a fantasia, an exuberant, ecstatic hymn to womanhood. Perhaps it is this that explains her more preposterous statements and a prose style that lurches from the purple to the polemical. But a fantasia is still no excuse for lapses of scientific good sense. To wit: “If breasts had something important to say, they would be much less variable and whimsical



Seat of life

Cleopatra, a woman who left her mark on history, but whose world was submerged by the Mediterranean, is revealed in *Cleopatra's Palace: In Search of a Legend* by Laura Foreman

(Random House, \$35). The book explores Cleopatra's era and how it came about with the aid of art and artefacts, such as this relief showing a woman in a birthing chair.

than they are," a statement that neglects the fact that secondary sexual characteristics are thought to vary a lot precisely because they convey important information to potential mates. Or, on why most people like other mammals better than insects, "we tend to like that which seems most like us, because resemblance implies genetic relatedness, and we like our genes; they have given us us". Forget that morphological similarity can evolve independently, or that few biologists would consider that kin selection — the theory that explains why organisms are more likely to cooperate with relatives than with strangers — can stretch to explain a preference for mammals over insects.

All this is a pity. It detracts from the places where her scientific explanations are good, and worse, it completely undermines the most interesting and important part of her book: her attack on the evolutionary psychology that Hales swallows whole.

Angier complains that evolutionary psychology — the current name for sociobiology as it is applied to people — is increasingly mooted as an explanation for everything from alleged differences in sex drive to why women tend to be paid less than men, despite a lack of convincing evidence for any of the claims. Indeed, even the most robust evidence — that there is a "universal" ideal for female beauty — has recently been dented. Thus, Angier wonders why, if human females develop breasts in adolescence to attract mates and if they have hidden ovulation, supposedly to encourage men to stick around, they have pregnancies that are so much more visible than those of other primates. Fair enough. Or consider the argument that is often made to explain patterns of marriage and divorce: that men prefer young women (because they are fertile) and polygyny (because that supposedly maximizes reproductive success) while women prefer men with more resources (because they make good fathers) and monogamy. Angier proposes an alternative: women are not usually economically independent enough to support a husband and family. If they were, some might well prefer handsome young toy-boys to old and ugly rich guys — nice personalities notwithstanding.

Of course, the real prediction that evolutionary psychology makes about female mating patterns is more subtle in any case. It is that women should prefer to maintain the appearance of monogamy, while practising polyandry. But poets like John Donne knew that four centuries ago:

*Though she were true, when you met her,
And last, till you write your letter,
Yet she
Will be
False, ere I come, to two, or three.* □

Olivia Judson is in the Department of Biology, Imperial College at Silwood Park, Ascot SL5 7PY, UK.



Chemical synthesis: Nernst's heat theorem was a logical extension of his multifaceted work.

A chemistry textbook hero

Walther Nernst and the Transition to Modern Physical Science

by Diana Kormos Barkan
Cambridge University Press: 1999. 288 pp.
£45, \$64.95

Charles Tanford and Jacqueline Reynolds

In his 1997 presidential address to the British Society for the History of Science, John Hedley Brooke deplored the inability of historians of science to appeal to a wider audience outside their own disciplinary confines. One explanation he offered is as follows: "In our scholarship we cannot help but become kill-joys. We dismiss legends, dissolve foundation myths, dilute the Eureka moments

and destroy the crucial experiments."

Well, those of us who are mere practising scientists can take heart from Diana Barkan's biography of Walther Nernst. Here is one book on the history of science that does none of these things. Admittedly, there is the obligatory (but in this case muted) criticism of scientists-cum-historians who have presented Nernst's heat theorem as the outcome of an explicit preoccupation with problems in chemical equilibria, rather than as a logical progression from his multifaceted, but related work. But his accomplishments are left intact, and even his legendary reputation as a difficult man is preserved, albeit with little supporting evidence. There is no mention of the famous quip, "you can't bury Nernst too often", ascribed to a colleague who was forced to attend all three successive interments of Nernst's remains.

Barkan characterizes her book as

“primarily a scientific biography” of Nernst, and it is just that: a long, carefully constructed trek through his published papers and associated studies from his contemporaries, as they apply to the scientific problems he investigated. Electrochemistry, thermodynamics, high- and low-temperature studies of the properties of materials, even his invention of an electrolytic lamp, are admirably woven together as the logical outcome of his own background and the scientific problems facing physical chemistry at the turn of the century. This was an exciting period in physical sciences, when atomism finally won the day against positivist diehards like Wilhelm Ostwald, and when radical theoretical changes appeared on the scene with Planck’s quantum theory and Einstein’s “*annus mirabilis*”.

All of the heroes and villains of our textbooks appear here, but, of course, they are not judged as such by the author. Ludwig Boltzmann, Nernst’s early mentor who sowed the seeds of atomism in his students; Svante Arrhenius, who first proposed the theory of electrolyte dissociation; Planck’s early sorties into the field of thermochemistry; Kamerlingh Onnes, discoverer of superconductivity (a phenomenon partially anticipated by Nernst during the period of his low-temperature studies); Fritz Haber, of the ammonia synthesis process, who, together with Nernst, participated in the German poison gas project during the First World War.

There are some shortcomings to the book, understandably so in such an ambitious undertaking. It is questionable whether Nernst was “a key figure in the transition to a modern, quantum theoretical physical science”, as the author claims. And with reference to thermodynamics, Nernst’s heat theorem is indeed often called “the third law of thermodynamics” in treatises on the subject, but just as often it is not given that distinction, because it isn’t logically comparable to the all-encompassing first and second laws. Barkan uses Helmholtz free energy to illustrate thermodynamicists’ need for a means to predict the conditions under which reactions occur spontaneously, but she never mentions the Gibbs free energy or the Gibbs chemical potential, which are more commonly used today, more useful and which had already been published by the period covered in the text.

Some of the key scientific figures, whose work is described in such meticulous detail, emerge as rather nebulous figures. The conflict between Planck and Nernst, for example, over the latter’s thermodynamic treatment of electromotive force is given an entire chapter, but not much emerges that gives insight into the characters involved. Only in the description of the 16-year battle by Arrhenius to prevent the Nobel Committee from awarding Nernst the coveted prize

do we get a glimpse of men rather than scientists. And here, when the last word has been read, one is moved to wonder if Nernst may have been more sinned against than sinner.

The final paragraph of Barkan’s book, summarizing the conclusions based on her research, states “The post-1895 revolution in the physical sciences happened individually, collectively, and in a variety of locales. It did not occur only because industry, the state, and the rising middle classes impinged and led to this transformation; or because science and scientists contributed to these social, political changes. Rather, it happened because particular experimental, theoretical, technological agendas were developed by specific people, research groups, countries, in unique ways because of distinctive, special circumstances.”

But although the prose is somewhat ponderous, could this possibly be an admission that individual scientists actually lead the way in scientific progress? And that individual scientists are prompted in what they aspire to by other individual scientists who preceded them?

In the final analysis, Barkan’s book provides an exceptionally convincing account of an era that is probably of greatest interest to physical chemists. Perhaps the narrow focus of the subject matter on a single person — one, moreover, whose place in the gallery of the greats is a matter for argument — throws a more concentrated light on the subject than would an historical tome of the period as a whole. We recommend this book, but, because it includes much that is controversial, we suggest that it be read in parallel with other recent books about the history of physical chemistry, for example Keith Laidler’s *The World of Physical Chemistry* (Oxford University Press), or John Servos’s *Physical Chemistry*

from Ostwald to Pauling: *The Making of a Science in America* (Princeton University Press), which have a different perspective, but cover much the same ground in scientific content. □

Charles Tanford and Jacqueline Reynolds are at Tarlwood, Back Lane, Easingwold, York YO61 3BG, UK.

Quantum behaviour

Introduction to Quantum Computation and Information

edited by Hoi-Kwong Lo, Sandu Popescu and Tim Spiller

World Scientific: 1998. 348 pp. \$52, £35

David P. DiVincenzo

Pure information, increasingly divorced from the physical products of farm and factory, is the lifeblood of modern civilization. Or so we are told. In our networked world, the digitized, packetized, encrypted creations appearing on our screens seem to have less and less to do with the ink and paper of which they used to be made. But the scientists who deliver these ethereal forms to us are increasingly concerned with how the processing and delivery of information is affected by its physical embodiment.

As the miniaturization of information progresses to its inevitable endpoint of one bit per atom, the unintuitive and paradoxical physics of the quantum will apparently increase our capacity to process and deliver information even further. The watershed theoretical developments of the past few years have shown that computations that are intractable on computers based on the laws of classical physics, such as the prime factorization of an integer, may become feasible if we use the quantum nature of the matter making up the computing device. Further-

New in paperback

Judging Science: Scientific Knowledge and the Federal Courts

by Kenneth R. Foster & Peter W. Huber
MIT Press, \$20, £13.95

“In its single-minded emphasis on making US courtrooms safe for science, the book overlooks crucial issues of law and public policy... And the tragic question of how justice can be rendered in a world of irreducibly uncertain knowledge is bypassed in silence.” Sheila Jasanoff, *Nature* 388, 635 (1997).

Bioarchaeology: Interpreting Behavior from the Human Skeleton

by Clark Spencer Larsen
Cambridge University Press, £24.95, \$39.95

Coral Reef Fishes: Caribbean, Indian

Ocean and Pacific Ocean including the Red Sea

by Ewald Lieske & Robert Myers
Princeton University Press, \$22.95

Queen Victoria’s Gene: Haemophilia and the Royal Family

by D. M. Potts & W. T. W. Potts
Sutton Publishing, £8.99, \$17.95

The Politics of the Earth: Environmental Discourses

by John S. Dryzek
Oxford University Press, £13.99

Dolphin Societies: Discoveries and Puzzles

edited by Karen Pryor & Kenneth S. Norris
University of California Press, \$22, £17.50

more, we can communicate more securely and efficiently if we pass quanta rather than bits over our networks.

These developments are admirably reviewed by the contributing authors to *Introduction to Quantum Computation and Information*. The book fills a gap between the turgid prose of the burgeoning research literature and the superficial accounts in the popular press.

Adriano Barenco lays out the basics of the factoring and searching quantum algorithms that captured the attention of much of the quantum-physics community. The first few chapters develop the basic language of quantum communication, and the unique features of the quantum state of separated quanta that we refer to as 'entanglement'. Michael Berry gives one of the most concise and correct definitions of quantum teleportation seen anywhere: "the dissolution of an object in one place in a way that enables it to be perfectly reconstituted elsewhere."

The theory of quantum cryptography and the extensive experiments and efforts at deployment in the field are also well surveyed. Other chapters deal with experimental efforts to implement quantum information processing in ion traps and in nuclear magnetic resonance spectroscopy.

Although the book's theoretical chapters will be useful for a long time, the experimental discussions may soon become obsolete in the face of entirely new approaches to quantum information processing, such as those provided by solid-state physics.

A striking feature of quantum information theory is that error correction is possible in quantum information processing. It was a great surprise that quantum error correction could be done even in principle, since the quantum phase is among the most fleeting, delicate and easily destroyed quantities in physics. The protocols that Andrew Steane and John Preskill helped to develop are not particularly simple or easy to understand, but their chapters are carefully constructed and very informative.

Preskill offers some of the most intriguing ideas in the book. There is a long-standing question of how the irreversible physics of black holes permits the quantum behaviour of our world to be apparently reversible to high precision. Preskill suggests that the true theory of gravity incorporates quantum error correction mechanisms that hide the irreversibility of gravity from our observations. If these ideas prove to be true, that most unphysical of quantities, pure information, may after all be capable of informing our understanding of the physical world in a most profound way. □

David P. DiVincenzo is at the IBM T. J. Watson Research Center, PO Box 218, Yorktown Heights, New York 10598, USA.

Science in culture



The Café Scientifique

www.aesops.force9.co.uk/cs/index.htm
Duncan Dallas

The Café Scientifique started last May in Leeds. I happened to read the obituary of Marc Sautet, the founder of the French 'Cafés Philosophiques', where people meet to discuss philosophy in cafés on Sunday mornings. It occurred to me that the British would never take this up, as they don't take philosophy seriously enough, but they might be prepared to discuss science — a subject they respect.

So in a local wine bar, I advertised an evening "where, for the price of a cup of coffee or a glass of wine, anyone can come to discuss the scientific ideas and developments which are changing our lives". On the first evening, with a speaker on Darwinism, I had no idea whether the audience might consist only of myself and a few friends.

In the event, between 40 and 50 people turned up, virtually the capacity of the café, and this has been the average size of the audience since then. There are a few regulars, but many new faces appear each week. We meet fortnightly, and invite a speaker to talk for about half an hour. Then there is a break to replenish our glasses, followed by questions and answers for an hour. Interest has been such that another Café Scientifique has been set up in Nottingham, with a further one planned for Newcastle.

Nowadays it appears necessary to fund university chairs to study how to make the public understand science. So why has this format worked so well?

There is no agenda, hidden or overt, to defend or sell science. If people don't like what they hear, they object forcefully. The subjects, or speakers, are picked because they are what people want to hear and they are often controversial. The audience sets the agenda, not the scientists. Not surprisingly, the biosciences feature heavily, but the café has also tackled chemistry, physics, maths and IT. The venue, a café-bar, is where the audience feel comfortable. The atmosphere is friendly and convivial, rather than academic and competitive. This is not a 'self-improving' audience, in the way that Victorian scientific societies arose. People don't just want to listen. They want to participate and be heard on equal terms with the scientists.

Science discovers café society, once the domain of poets and writers.

It is also, as it turns out, an income-generating idea for the wine bar, so there are no start-up or organizational costs. Indeed, we pay the speakers' expenses by passing a hat, 'le chapeau scientifique', round the audience. Since the Café Scientifique works in a suburb of Leeds (not even the university suburb), it could work anywhere.

Although I didn't know it, at around the same time in France, other Cafés Scientifiques were appearing, no doubt also inspired by the *Cafés Philosophiques*, so it looks as though it is an idea of its time, possibly even a movement.

There may well be something of interest happening here, something of real importance to science, and I use an analogy from medicine. In the late 1970s, medicine was a successful, authoritarian profession, rather like much science today. By and large, patients gratefully accepted any treatment from doctors. During the 1980s all of this changed — the fitness movement started, alternative medicine became established, funding crises proliferated, AIDS became a problem, self-help groups and ethics committees grew up, anxiolytic drugs were found to be addictive, and so on. Today, medicine is a battleground of politics, economics and philosophy. This has not stopped the progress of orthodox medicine, but doctors now work in a very different environment, subject to different constraints and conditions.

Since the end of the Cold War, science has come to the centre of the agenda in environmental politics (global warming), economics (BSE), consumer choice (GMOs) and medicine (genetics).

Under these circumstances science can no longer be taken on its own terms. Many new ways of addressing science will be invented, not all of them rational or favourable. The Café Scientifique is just a harmless straw in what might become a cold wind. What is critical is that scientists don't repeat the early mistakes made by doctors, who often buried their heads in the sand and hoped the new attitudes would go away. Scientists can no longer define the terms of the debates about science. □

Duncan Dallas is at XYTV, *The Old Stables*, 7 Regent Street, Leeds LS7 4PE, UK.

Supersonic winds in Jupiter's aurorae

Daniel Rego*, Nicholas Achilleos*, Tom Stallard*,
Steve Miller*, Renée Prangé†, Michele Dougherty‡
& Robert D. Joseph§

* Department of Physics and Astronomy, University College London, Gower Street,
London WC1E 6BT, UK

† Institut d'Astrophysique Spatiale, UMR-CNRS 120, Batiment 121,
Université de Paris XI, 91405 Orsay Cedex, France

‡ Space and Atmospheric Physics, Imperial College, London SW7 2BZ, UK

§ Institute for Astronomy, University of Hawaii, Woodlawn Drive, Honolulu,
Hawaii 96822, USA

Jupiter has a giant magnetosphere that is coupled to the planet's upper atmosphere; as the planet rotates, its magnetic field drags a dense ionized equatorial sheet of plasma, which must interact with the upper atmosphere. Jupiter's aurorae are much more powerful^{1,2} than the Earth's, and cause significant local heating of the upper atmosphere. Auroral electrojets—ion winds that race around Jupiter's auroral ovals—play a key role in theoretical models of how Jupiter's rotational energy is transferred to the plasma sheet^{3,4} and how winds may transport energy from auroral heating to lower latitudes^{5–7}. But there has hitherto been no direct observational evidence for the existence of such electrojets. Here we report observations of electrojets that have winds approaching or in excess of the local speed of sound. The energy produced by these electrojets could heat the whole upper atmosphere, if the auroral regions couple efficiently with the rest of the planet.

In standard models of the jovian magnetic field, the plasma

sheet—which is several jovian radii, R_J , thick—extends from $5R_J$ and co-rotates with the planet out to $\sim 20\text{--}30R_J$, where co-rotation breaks down⁸; many of the energetic particles which produce aurorae originate from here. (Note that $5R_J$ is inside the orbit of Io, which is located at $5.9R_J$.) According to Hill³, co-rotation is enforced by a large-scale current system which flows radially outwards in the plasma sheet, and closes by Birkeland (field-aligned) currents (upward from Jupiter nearer to Io's orbit, and downward from the more distant magnetosphere), and largely meridional Pedersen currents in the ionosphere. There, the magnetic and electric fields should result in anti-corotational (clockwise) ion winds ($\mathbf{E} \times \mathbf{B}$ Hall drift)—the electrojets. These electrojets flow along the oval(s) mapped onto the planet by the aurorally active magnetic field lines. The electrojets are also expected to accelerate (through friction) the ambient neutral gas⁷ and—together with other processes—to generate fast thermospheric winds needed to redistribute part of the auroral heating⁶.

The magnetospheric Birkeland currents required by Hill's mechanisms have indeed been detected⁹, but there was only very indirect evidence for the mechanisms in the jovian ionosphere/thermosphere^{10,11}. Our direct evidence for auroral electrojets is contained in high-resolution infrared spectra obtained using the NASA Infrared Telescope Facility (IRTF) on Mauna Kea, Hawaii. These spectra were taken at $3.953\text{ }\mu\text{m}$, a wavelength sensitive to the H_3^+ molecular ion, on the night of 8 August 1997, using the CSHELL spectrometer. Our observations appear to have encompassed an 'auroral event', perhaps similar to those known in ultraviolet data^{9,12–15}.

CSHELL operates both as a long-slit spectrometer and as a direct imager with a pixel scale of 0.2 arcsec (corresponding to a linear distance at Jupiter of $\sim 680\text{ km}$). Figure 1 shows a sequence of circular variable filter (CVF) images of the northern polar auroral region of Jupiter, taken on 8 August 1997. Unlike data taken at

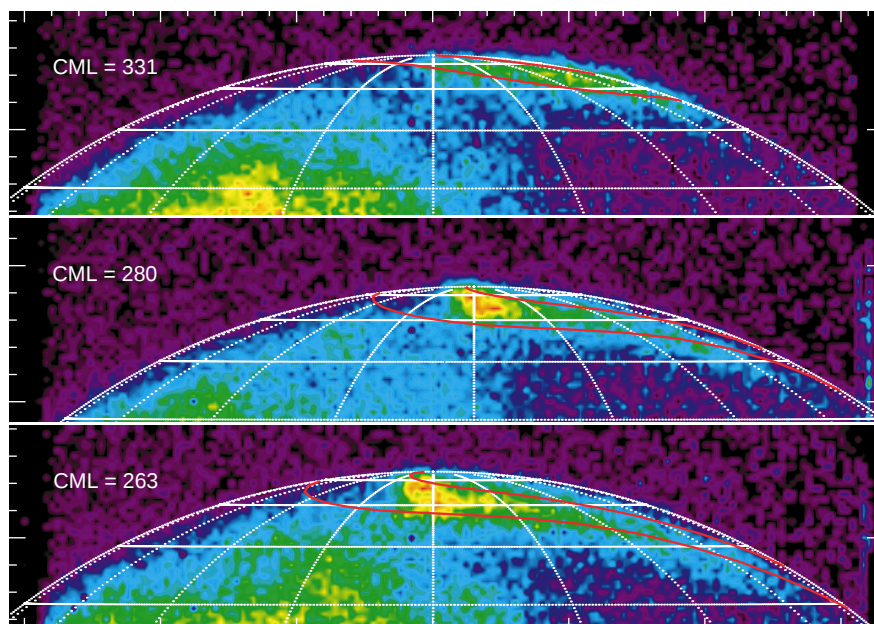


Figure 1 Series of CVF (1.4% spectral resolution) images of the northern auroral region of Jupiter. Images were obtained on 8 August 1997: bottom panel, at 08:54 UT; middle, at 09:22 UT; top, at 10:46 UT. Central wavelength, $3.953\text{ }\mu\text{m}$; exposure time, 5 s; pixel scale, 0.2 arcsec ; seeing, $\sim 0.8\text{ arcsec}$; tracking accuracy, $\sim 0.1''\text{ min}^{-1}$; CML, central meridian longitude. A planetary grid (white dotted lines: longitudinal divisions 18° ; latitudinal divisions 10°), which includes the $30R_J$, (more poleward) and $5.9R_J$, (more equatorward) field line footprints from the most recent VIP4 jovian magnetic field model², has been fitted to the limb. Each division on the axes is 1 arcsec , the width of the CSHELL slit used for the spectral images. The bright emission lying along the magnetic field line footprints is due to the $\text{H}_3^+ \nu_2(1, 0^-)$

line at $3.953\text{ }\mu\text{m}$; incident solar radiation at high latitudes is absorbed by jovian stratospheric methane. The bright emission seen on the body of the planet at lower latitudes is due to reflected solar infrared radiation: the shorter CH_4 column is insufficient to absorb all the incident radiation here. The central wavelength is not constant across the array when CSHELL is used in CVF mode, and this limits the use that can be made of our images to visualizing the overall structure, rather than investigating the detailed morphology. The images show that both arcs are within the visible limb of the planet, and correspond approximately to the magnetic footprints shown.

similar central meridian longitudes (CMLs) a year previously, which showed a single auroral arc, these images showed a double partial arc close to the footprints of the $30R_J$ (poleward) and $5.9R_J$ (equatorward) magnetic field lines, respectively (Fig. 1 bottom and middle panels). Our observations of the 'double arc' event lasted about half an hour; two hours after our initial observation only a single auroral arc showed (Fig. 1 top panel).

Used in echelle spectrometer mode, each wavelength pixel corresponds to a dispersion of 2.9 km s^{-1} ; with the slit width of 1 arcsec (that is, 5 pixels) used in this study, CSHELL has a spectral resolving power of 20,700 (14.4 km s^{-1}) for a source which fills the slit. For the spectral images reported here (Fig. 2 and Table 1), the slit was aligned along the central meridian from pole to equator in order to obtain latitudinal profiles of the planet's H_3^+ emission; as the CSHELL aperture spans only 30 arcsec, northern and southern hemispheres were observed separately. Spectral images were obtained in pairs of 60-s integrations on the planet (and 60 s on the sky, for background subtraction), and a four-minute total elapse time. Any necessary corrections to the wavelength scale were made

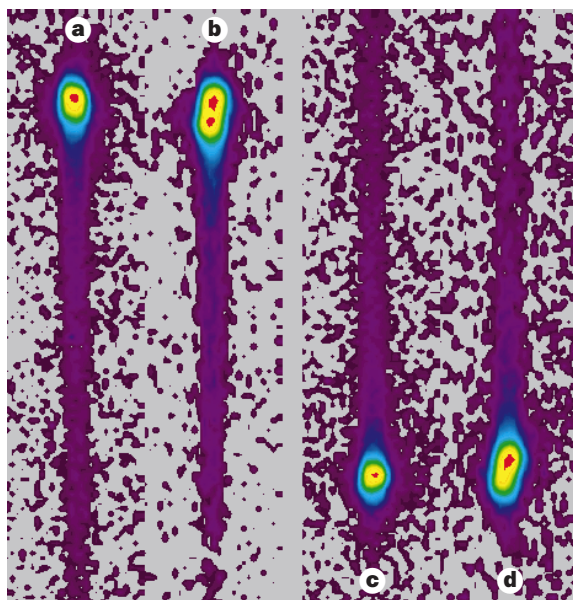


Figure 2 Comparison of CSHELL spectral images from 14 and 16 July 1996, and 8 August 1997. These show the 1997 'two-peak' structure in the north and the 'peak-tail' structure in the south, not visible in the 1996 data. **a**, 16 July 1996, 09:56 UT. CML = 248° , northern hemisphere. **b**, 8 August 1997, 08:55 UT. CML = 264° , northern hemisphere. **c**, 14 July 1996, 08:46 UT. CML = 264° , southern hemisphere. **d**, 8 August 1997, 09:05 UT. CML = 270° , southern hemisphere. The CSHELL 30"-long slit is set along the CML and the vertical axis indicates the position on the planet; north is at the top in the figure. The spectral images cover Jupiter from the poles—where the bright auroral emissions are located—to the equatorial regions, and each hemisphere is measured separately. The horizontal direction indicates wavelength (increasing left-to-right) for each of the four (**a–d**) spectral images shown. The images have been flat-fielded to allow for the nonlinear response of the array, and bad pixels have been removed. The central wavelength for each spectral image is $3.95301 \mu\text{m}$, corresponding to the rest wavelength of the $\text{H}_3^+ \nu_2 \text{ Q}(1,0)$ line. The wavelength has been corrected for any slight non-alignment of the slit on the array and for variation in the wavelength calibration along the slit (total effect <1 pixel for the entire 150-pixel, 30-arcsec, slit-length), using atmospheric emission lines at $3.95251 \mu\text{m}$ and $3.95416 \mu\text{m}$; in the case of our data, no discernible systematic wavelength shift was observed. The other "event" spectral images, listed in Table 1, show the same wavelength structure as **b**, for the northern hemisphere, and **d**, for the south. Post-event spectral images have the same structure as **a**, for the north, and **c**, for the south; that is, they show neither the 'double-peak' nor the 'arc-tail' structure, nor any large or systematic wavelength shifts.

Table 1 Derived line-of-sight velocity shifts, 8 August 1997

Time (UT)	CML (deg.)	IP (km s^{-1})	OP (km s^{-1})	Difference (km s^{-1})
During event				
08:50	260 N	1.0	-1.8	-2.8
08:55	265 N	0.9	-1.3	-2.2
09:05	270 S	-0.8	2.5	3.3
09:10	274 S	-1.0	2.0	3.0
09:18	278 N	0.0	-2.3	-2.3
09:24	282 N	0.3	-2.1	-2.4
09:33	287 S	-0.7	2.0	2.7
09:38	290 S	-1.0	1.5	2.5
N average		0.6	-1.9	-2.5
S average		-0.9	2.0	2.9
After event				
10:38	327 N	0.2	-1.1	-1.3
10:42	333 N	0.1	-1.1	-1.1
10:55	337 S	-0.4	0.4	0.8
11:00	341 S	-0.6	0.2	0.8
11:12	347 N	0.6	-0.6	-1.2
11:16	350 N	0.1	-0.8	-0.9
11:24	355 S	-0.4	-1.2	-0.8
11:32	359 S	-0.2	-0.6	-0.4
N average		0.2	-0.8	-0.8
S average		-0.4	-0.3	0.1

Table shows derived wavelength (velocity) shifts of the inner (equatorward) peak (IP) and outer (poleward) peak/tail (OP) for pairs of spectra, and the difference (OP – IP) between them, for the jovian auroral regions. The CML is given in System III longitude, and the letter 'N' or 'S' denotes measurement of either the northern or southern auroral region. Pairs of spectra, obtained within 4 min elapsed time of one another, have been combined to improve statistics; UT times refer to the observation mid-time. Errors in the IP, OP and difference velocity shifts for individual spectra are 0.5 km s^{-1} and for pairs of spectral images are $\pm 0.3 \text{ km s}^{-1}$ (see main text); the averages are subject to errors of $\pm 0.2 \text{ km s}^{-1}$. After the event, the outer structures were not seen, and the velocity values refer to the maximum measured poleward of the main peak; this often refers to very low intensity emission, with signal-to-noise (S/N) < 3 , as against the event velocities which all refer to high intensity emission with $S/N > 10$.

by reference to sky emission lines whose wavelengths lay close to the H_3^+ line (see legends to Figs 2 and 3).

Figure 1 shows that at the time of our observations in the north, the positioning of the CSHELL slit meant that it cut both the partial arcs, and that it crossed the poleward arc, in particular, very close to the rising (eastern) ansa, the turning point of the auroral arc. The resulting H_3^+ latitudinal profiles showed a 'double peak' structure (Fig. 2b) in comparison with the single peak seen in 1996 (Fig. 2a). In addition, the wavelength of the outer (poleward) peak was red-shifted by ~ 1 pixel, in comparison with the inner (equatorward) peak. In the south, the images showed an arc plus diffuse emission and the H_3^+ latitudinal profiles an 'arc-tail' structure (Fig. 2d), not seen in 1996 (Fig. 2c). The images showed that the slit crossed the auroral arc near the setting (western) ansa; the southern 'tail' was blue-shifted by ~ 1 pixel. This behaviour was repeated in eight pairs of spectra—four in the north and four in the south—taken between 8:50 and 9:38 UT, which appear to have spanned our 'auroral event', but was absent in spectra taken after that time (Table 1).

To measure the spectral shift in the $3.953\text{-}\mu\text{m}$ H_3^+ line in the auroral regions, we fitted the intensity profile of the spectrum for each row of pixels with a gaussian function in which the amplitude, half-width, central wavelength, and background were free parameters. To calibrate the wavelength in the jovian frame of reference, the H_3^+ line emitted on the body of the planet was measured on the same spectral images (see Fig. 2 legend). This planetary reference spectrum was obtained by summing the spectra along ~ 70 rows on the array (that is, ~ 14 arcsec) equatorward from the sub-auroral regions. Spectral shifts of the auroral emission peaks for those spectral images which exhibited the double-peak or peak-tail structure are given in Table 1, along with the results for the spectral images obtained just after the double-peaked structures disappeared. For each spectrum with a signal-to-noise ratio greater than 10, that is, for all of the spectra in the auroral and near-auroral regions, the fitting procedure gives peak positions to an accuracy of ± 0.1 pixels or $\pm 0.3 \text{ km s}^{-1}$. As the error is the same in fitting the profile on the body of the planet, the uncertainty in determining the

shifts between either peak and the planetary rest wavelength is approximately $\pm 0.5 \text{ km s}^{-1}$. Typical velocity profiles for the auroral regions are shown in Fig. 3.

For the period when the two-peaked structure was present, Table 1 shows that in the northern auroral zone, the poleward peak is seen to have a mean redshift, relative to the peak 1 arcsec towards the equator, of $-2.5 \pm 0.2 \text{ km s}^{-1}$. For the southern auroral zone, the relative blueshift between the equatorward peak and the poleward tail is $+2.9 \pm 0.2 \text{ km s}^{-1}$. In each hemisphere, part of this relative shift is due to shifting of the equatorward peaks ($+0.6 \text{ km s}^{-1}$ in the north, -0.9 km s^{-1} in the south). After the two-arc structure disappeared, spectra from the limb of the planet showed shifts of $-1 \pm 0.2 \text{ km s}^{-1}$ in the north and $+0.1 \pm 0.2 \text{ km s}^{-1}$ in the south, with respect to the single auroral peak—effectively zero spectral shift in the south, and a small residual shift in the north.

Placing a point source asymmetrically in the slit could produce an

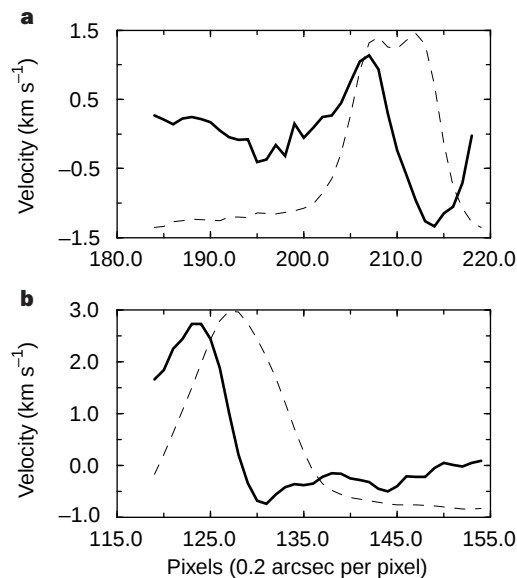


Figure 3 Line-of-sight velocity shifts and intensity profiles for the spectral images shown in Fig. 1 for the $\text{H}_3^+ \nu_2 \text{Q}(1,1)$ line. Planetary south is on the left and north on the right of the figure. **a**, Northern hemisphere CML 270°; **b**, southern hemisphere CML 264°. Solid lines show velocities expressed as shifts from the planetary rest wavelength in km s^{-1} , determined from the wavelength of the non-auroral emission from the main body of the planet. Systematic and fitting errors in our peak positions were determined by fitting the sky emission lines to a gaussian profile; these were shown to be $\pm 0.3 \text{ km s}^{-1}$ for five-pixel bins for the (relatively weak) sky lines on top of the bright sky continuum. The $\text{H}_3^+ \nu_2 \text{Q}(1,0^-)$ line was fitted for the sky-subtracted frames, so that the departure from gaussian profile caused by the adjoining sky lines and non-constant sky continuum was minimized. The emission from the main body of the planet, from the equator to the sub-auroral latitudes, is unaffected by Doppler shifts due to east-west winds, as the slit was positioned along the CML, giving a zero line-of-sight component. North-south wind speeds in the temperate to sub-auroral zones are $< 100 \text{ m s}^{-1}$ (refs 14 and 19), and vertical velocities are even less. Thus this body-of-planet emission enables us to derive a planetary rest wavelength from which the shifts of the auroral peaks/tails can be measured. Taking into account the error in fixing the planetary rest wavelength, each individual pixel velocity is subject to an error of $\pm 0.5 \text{ km s}^{-1}$. The overplotted dashed line is the corresponding intensity (arbitrary units). In both hemispheres, the outer intensity peak (northern hemisphere) or the tail (southern hemisphere), which corresponds approximately to the locus of the $30R_J$ field line, is associated with the greatest velocity shift; the inner peak, corresponding approximately to the locus of the $5.9R_J$ (Io orbit) field line, is associated with a slight velocity shift in the opposite sense to the outer peak. (A detailed description of our data analysis is obtainable by anonymous ftp at jonny.phys.ucl.ac.uk; change directory to the pub/aurwinds/referees directory and download the .ps files.)

apparent wavelength shift. But a detailed examination of this effect (not shown), making use of both images and spectra, ruled it out for the following reasons. First, the images showed emission extended along the two arcs in the east–west direction for considerably more than 1 arcsec, the width of the CSHELL slit; and, taking the northern images shown in Fig. 1 for instance, there was some intensity variation in the east–west direction. But a slit positioning which put the brightest part of the poleward arc on one side of the slit and that of the lower-latitude arc on the other side would have actually produced a relative blueshift for the outer peak, in contrast to what was observed. Second, the same spectral shifts were obtained for four spectral images in each of the northern and southern hemispheres, despite offsetting the telescope several times, over a period when Jupiter rotated much more than one slit width. The likelihood of producing such consistent wavelength shifts as a result of an essentially randomized positioning of point sources in the slit seemed negligible.

Thus our observations are consistent with H_3^+ ions flowing in a clockwise, anti-corotational, direction along the $\sim 30R_J$ auroral ovals in both hemispheres, as predicted by Hill³ to enforce the co-rotation of the magnetospheric plasmasheet, and—with lower velocities—near the $\sim 5.9R_J$ auroral ovals in the opposite direction. This is more unexpected but Morgan *et al.*¹⁶ have proposed currents flowing radially towards Jupiter in the plasma sheet for about $0.2R_J$ inward of Io's orbit (a dense region called the "ribbon") to explain "auroral hiss" measured by Voyager. Such currents would drive anticlockwise ionospheric winds, and may therefore explain the shifts which appear to be associated with our inner ($\sim 5.9R_J$) peaks.

The speed of sound in the thermosphere in the region where H_3^+ is formed is 2.6 km s^{-1} (ref. 17). As we detect only $\sim 70\%$ of the wind speed along the $\sim 30R_J$ ovals, our velocities in Table 1, corrected for the line of sight, indicate that the ion flows are close to the speed of sound or supersonic. The highest corrected velocities are $3.3 \pm 0.4 \text{ km s}^{-1}$ for the north outer peak and $3.6 \pm 0.4 \text{ km s}^{-1}$ for the south 'tail', with averages of $2.7 \pm 0.3 \text{ km s}^{-1}$ and $2.9 \pm 0.3 \text{ km s}^{-1}$, respectively. This detection of near-to-supersonic electrojets provides direct evidence that high-velocity winds do exist in the jovian thermosphere/ionosphere, and is powerful confirmation of our present understanding of the jovian magnetosphere and the ionospheric current systems that it produces. □

Received 13 October 1998; accepted 23 March 1999.

1. Livngood, T. A., Moos, W. A., Ballester, G. E. & Prangé, R. Jovian ultraviolet auroral activity, 1981–1991. *Icarus* **97**, 26–45 (1992).
2. Lam, H. A. *et al.* A baseline spectroscopic study of the infrared auroras of Jupiter. *Icarus* **127**, 379–393 (1997).
3. Hill, T. W. Inertial limit on corotation. *J. Geophys. Res.* **84**, 6554–6558 (1979).
4. Hill, T. W., Dessler, A. J. & Goertz, C. K. in *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) 353–394 (Cambridge Univ. Press, 1983).
5. Atreya, S. K. *Atmospheres and Ionospheres of the Outer Planets and their Satellites* 107–134 (Springer, Heidelberg, 1986).
6. Waite, J. H. Jr *et al.* Electron precipitation and related aeronomy of the jovian thermosphere and ionosphere. *J. Geophys. Res.* **88**, 6143–6163 (1983).
7. Somneria, J., Ben Jaffel, L. & Prangé, R. On the existence of supersonic jets in the upper atmosphere of Jupiter. *Icarus* **119**, 2–25 (1995).
8. Connerney, J. E. P. Magnetic fields of the outer planets. *J. Geophys. Res.* **98**, 18659–18679 (1993).
9. Dougherty, M., Dunlop, M., Prangé, R. & Rego, D. Correspondence between field-aligned currents observed by Ulysses and Hubble Space Telescope auroral emission. *Planet. Space Sci.* **46**, 531–540 (1998).
10. Prangé, R. *et al.* Detection of self-reversed Lyman- α lines from the jovian aurorae with the Hubble Space Telescope. *Astrophys. J.* **484**, L169–L173 (1997).
11. Miller, S. *et al.* Mid-to-low latitude H_3^+ emission from Jupiter. *Icarus* **130**, 57–67 (1997).
12. Prangé, R. *et al.* Correlated variation of u.v. and radio emission during an outstanding auroral event. *J. Geophys. Res.* **98**, 18779–18792 (1993).
13. Gérard, J. C. *et al.* A remarkable auroral event observed in the ultraviolet with the Hubble Space Telescope. *Science* **266**, 1675–1678 (1994).
14. Ballester, G. E. *et al.* Time-resolved observations of Jupiter's far-ultraviolet aurora. *Science* **274**, 409–413 (1996).
15. Prangé, R. & Livngood, T. A. in *Ultraviolet Astrophysics Beyond the IUE Final Archive* (SP-413, ESA, Paris, 1998).
16. Morgan, D. D., Gurnett, D. A., Kurth, W. S. & Bagenal, F. The source of jovian auroral hiss observed by Voyager 1. *J. Geophys. Res.* **99**, 21213–21224 (1994).
17. Achilleos, N. *et al.* JIM: a time-dependent, three-dimensional model of Jupiter's thermosphere and ionosphere. *J. Geophys. Res. (Planets)* **E103**, 20089–20112 (1998).
18. Clarke, J. T. *et al.* HST far-ultraviolet imaging of Jupiter during the impacts of Comet Shoemaker-Levy 9. *Science* **267**, 1302–1307 (1995).

19. Dinelli, B. M. *et al.* UKIRT observations of the impact and consequences of Comet Shoemaker-Levy 9 on Jupiter. *Icarus* **126**, 107–125 (1997).

Acknowledgements. N.A. was a visiting astronomer on the NASA Infrared Telescope Facility, operated by the Institute for Astronomy, University of Hawaii, on behalf of NASA. We thank staff at the NASA IRTF for assistance. In particular, we thank J. Rayner for help in dealing with issues related to CSHELL. This work was supported by the UK Particle Physics and Astronomy Research Council.

Correspondence and requests for materials should be addressed to S.M. (e-mail: s.miller@ucl.ac.uk).

Efficient fault-tolerant quantum computing

Andrew M. Steane

Department of Atomic and Laser Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK

Quantum computing¹—the processing of information according to the fundamental laws of physics—offers a means to solve efficiently a small but significant set of classically intractable problems. Quantum computers are based on the controlled manipulation of entangled quantum states, which are extremely sensitive to noise and imprecision; active correction of errors must therefore be implemented without causing loss of coherence. Quantum error-correction theory^{2–9} has made great progress in this regard, by predicting error-correcting ‘codeword’ quantum states. But the coding is inefficient and requires many quantum bits^{10–12}, which results in physically unwieldy fault-tolerant quantum circuits^{10–18}. Here I report a general technique for circumventing the trade-off between the achieved noise tolerance and the scale-up in computer size that is required to realize the error correction. I adapt the recovery operation (the process by which noise is suppressed through error detection and correction) to simultaneously correct errors and perform a useful measurement that drives the computation. The result is that a quantum computer need be only an order of magnitude larger than the logic device contained within it. For example, the physical scale-up factor^{10,11} required to factorize a thousand-digit number is reduced from 1,500 to 22, while preserving the original tolerated gate error rate (10^{-5}) and memory noise per bit (10^{-7}). The difficulty of realizing a useful quantum computer is therefore significantly reduced.

The inefficient quantum error correcting codes (QECCs) mentioned above are not the only way to perform this function; more efficient QECCs exist^{4–6,19–21}, and there exist fault-tolerant methods to extract syndromes for these codes^{10,15,16}. Knill¹⁷ discussed ways to find logical operations on general codes and more recently Gottesman^{18,22} derived a complete set of fault-tolerant operations which can work with efficient QECCs. However, these methods are not in themselves sufficient, because the greater complexity of the operations lowers the error tolerance, offsetting the gain in coding efficiency.

This deficiency is overcome here by one central concept and a combination of several subsidiary ones. The central concept is to adapt the recovery operation so that while extracting an error syndrome, a desired measurement on the logical state of the computer is also made. This essentially reduces to zero the cost of making such measurements, and they may be used to construct useful logical operations such as swaps (via teleportation). The subsidiary ideas concern the best choice of QECC, the most efficient use of ancillary qubits when extracting syndromes, and the fast preparation of these ancillary bits.

I will implement quantum networks by means of the universal set of logical gates $\{H, P, {}^CX, T\}$ (ref. 13). Here H is the Hadamard operation, $P = |0\rangle\langle 0| + i|1\rangle\langle 1|$, CX is controlled-NOT and $T \equiv {}^{CC}X$ is the Toffoli gate. Using T and measurement permits general single-

bit rotations¹¹. Any useful quantum computation must make significant use of T , otherwise it could be efficiently simulated classically^{18,23}. For example, Shor’s algorithm is limited by the Toffoli gates required to evaluate modular exponentials^{11,24,25}.

The Calderbank Shor Steane (CSS) quantum codes are those whose stabilizer generators separate into $\bar{X}$ and $\bar{Z}$ parts^{19,21}. I restrict attention to these codes, because they permit a larger set of easy-to-implement fault-tolerant operations, and their coding rate k/n can be close to that of the best codes. The codewords are

$$|u\rangle_L = \sum_{x \in C_0} |x + u \cdot \bar{D}\rangle \quad (1)$$

where u is a k -bit binary word, and the right-hand side is a superposition of n -bit product states. C_0 is a linear classical code and $\bar{D}$ is a $k \times n$ binary matrix of coset leaders (the tilde will indicate symbols which refer to binary matrices). Let $\bar{H}$ be the generator matrix of C_0 .

I define an operation to be ‘legitimate’ if it maps the encoded Hilbert space onto itself. An operation is fault-tolerant if it does not cause errors in one physical qubit to propagate to two or more qubits in any one block. Bitwise application of a two-bit operator is defined to mean the operator is applied once to each pair of corresponding bits in two blocks. Legitimate bitwise operations are fault-tolerant.

Of the CSS codes, those for which C_0 is doubly even (all words have weight 0 mod 4) and $\bar{D}\bar{D}^T = \bar{I}$ are especially useful, as it can be shown^{18,26} that in this case bitwise ${}^CX, H$ (hence also CZ where $Z = HXH$) and P are legitimate and perform the corresponding logical operations on whole blocks at a time. To operate on individual encoded qubits, logical qubits are switched into otherwise empty blocks, whole-block operations are applied, and the qubits are then switched back. Finally, the Toffoli gate is needed, for which I use inter-block switching together with Shor’s¹³ implementation. The Shor technique works for any codes in which bitwise CZ acts as blockwise logical CZ , so it is valid for our chosen codes.

Gottesman¹⁸ proposed switching and swapping using teleportation. This and further methods²² make much use of the ability to measure $\bar{X}$ or $\bar{Z}$ operators fault tolerantly (here I use the overbar to indicate operations in the logical Hilbert space). A method to perform such measurements has been proposed¹⁵, based on the preparation of verified ‘cat’ states $|000\dots 0\rangle + |111\dots 1\rangle$. However, the preparation and verification of these states involves many elementary gates, and the measurement must be repeated to ensure reliability. These operations take a considerable number of elementary gates and time steps, during which errors accumulate. This significantly reduces the tolerance on error rates in the computer. I now introduce an important ‘trick’ to circumvent this problem: I show how, for any stabilizer code, measurement of any operator $\bar{X}_u$ or $\bar{Z}_u$ can be performed at no cost by merging it with the recovery operation.

Here, the subscript u indicates the logical bits to which the operator is applied. I will implement fault-tolerant recovery using the method of Steane^{10,16}, which is based on preparing a $2n$ -bit ancilla in a superposition of 2^{n+k} product states which satisfy the parity checks in the stabilizer. The measurement technique which underlies the method is illustrated for a CSS code in Fig. 1. In order to measure $\bar{X}_{010}$ in this example I prepare an ancilla in $|000\rangle_L + |010\rangle_L$ and operate bitwise CX , then Hadamard-transform the ancilla and measure it. This permits one to learn simultaneously the result of measuring $\bar{X}_{010}$ on the logical qubits, and the syndrome for Z errors, which can then be corrected (the whole network is repeated as necessary; see below). Replacing CX by CZ , a measurement of $\bar{Z}_u$ can be accomplished while learning simultaneously the syndrome for X errors. The structure of CSS codes permits the $2n$ -bit ancilla to consist of two separate blocks of n bits, which is why Fig. 1 only shows an n -bit ancilla. For general stabilizer codes the method is essentially the same but does not have such an elegant

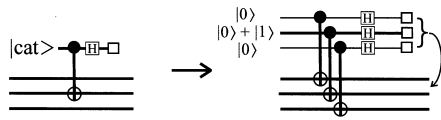


Figure 1 Method to measure $\bar{X}_{010}$ while simultaneously extracting the syndrome for Z errors. The left-hand network is that proposed in ref. 15; it is replaced by the right-hand network. The horizontal lines represent logical qubits, the boxes represent the measurement of the ancilla, from which a syndrome and the required $\bar{X}_{010}$ measurement can be deduced. The final arrow represents correction of the deduced Z errors. The necessary repetition discussed in the text is not shown.

expression in terms of logical qubit states.

A standard recovery, without measurement of any observable other than the syndrome, involves the preparation of $|0\rangle_L$. To prepare $|0\rangle_L + |u\rangle_L$ from $|0\rangle_L$ I simply add a one-bit Hadamard transform and C 's to target bits at the non-zero coordinates in $u \cdot \bar{D}$. This is only slightly more complicated than preparation of $|0\rangle_L$ because the number of rows in $\bar{H}$, which gives the network to construct $|0\rangle_L$, is much larger than 1 for powerful QECCs. Furthermore, as $|0\rangle_L + |u\rangle_L$ satisfies fewer parity checks than $|0\rangle_L$, the verification of the prepared state is quicker. Hence the statement "at no cost" in the above 'trick' is justified.

Figure 2 gives the implementation of the Toffoli gate described by Preskill¹¹ based on Shor's ideas. Overall, the Toffoli network requires about eight recoveries.

Next I identify useful quantum codes. The condition for a Bose Chaudhuri Hocquenghem (BCH) classical code²⁷ to contain its dual is known^{21,28}. I conjecture that the dual of a BCH code of length $2^m - 1$ is doubly even whenever the code contains its dual. I have confirmed this for $n \leq 127$ by examining the parity check matrices, and it is provable for $m > 3$ for the double- and triple-error correcting cases²⁷. Hence there is a large class of useful codes, containing $[[n, k, d]] = [[31, 11, 5]]$, $[[31, 1, 7]]$, $[[63, 39, 5]]$, $[[63, 27, 7]]$, $[[127, 85, 7]]$, $[[127, 43, 13]]$ and $[[127, 29, 15]]$, plus $[[255, 143, 15]]$ by conjecture.

I now estimate the amount of noise which can be tolerated by a quantum computer using the methods discussed. The analysis is similar to one previously described¹⁰ but with some new features. The treatment of noise has been discussed^{2,6-9}; it is a continuous process affecting all qubits and all gates, but each syndrome measurement collapses the system onto a state which differs from the noise-free state by a discrete error operator. I calculate the probability that this operator is uncorrectable by the QECC.

A computation involving K logical qubits will be carried out using K/k blocks to store the quantum information, plus a further three blocks which only store one logical qubit at a time. These act as the 'accumulator' for the Toffoli network, and for other logical gates as necessary.

In order to extract syndromes, I prepare, in parallel, four ancillary blocks for every data block or accumulator block in the computer. A single complete syndrome is extracted, using two of the prepared ancillas for each block. Wherever this first syndrome indicates no errors, it is accepted even though it has a small chance of being the wrong syndrome²⁹. Any errors in such blocks will with high probability be corrected by the next recovery. The remaining blocks are corrected, but only after their syndrome has been extracted a further $O(d)$ times, and the best estimate syndrome is taken. The ancillas required have already been prepared: they are the remaining $2(3 + K/k)$ ancillas which were not used for the first syndrome. These are sufficient in number when the probability of obtaining a non-zero syndrome is less than $O(1/d)$, which it is for the cases I will consider.

Each ancilla uses n qubits to store the prepared state, plus one for verification, therefore the total number of physical qubits in the

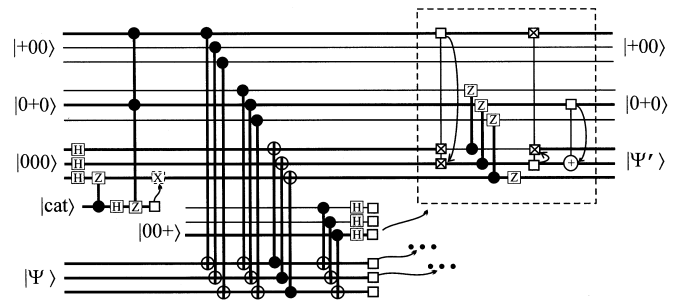


Figure 2 Fault-tolerant Toffoli gate on three qubits in the same block. If the data block at the bottom is initially in the state $|\psi\rangle_L$, then the third ancillary block ends up in the state $\bar{T}|\psi\rangle_L$. The method is illustrated for a QECC encoding three qubits per block, but applies essentially unchanged for all CSS codes based on self-dual classical codes. Horizontal lines represent logical qubits, and the notation $|00+\rangle$ means $(|000\rangle_L + |001\rangle_L)/\sqrt{2}$, and so on. The fourth ancilla is an example of Fig. 1 and the method proposed in the text; it is used to measure $\bar{X}_{001}$ on the data block (and extract a syndrome). The operations in the dashed box are only carried out if this measurement yields a 1. The figure does not show further operations which are required if the other measurements yield 1s (see ref. 11 for details) or the repetition of the initial "measurement via cat" which prepares the ancillas. During its preparation, the cat is verified against X errors in order that the CZ does not propagate uncorrectable errors into the third ancilla. However, the cat has a chance of acquiring a Z error, which makes the measurement fail with probability linear in the error rates. We therefore repeat this part of the network, using further n -bit cats prepared in parallel, and a majority vote is taken after $O(d)$ repetitions. The three-qubit operations in the dashed box are teleportations in the direction of the arrow, and each can be accomplished with two measurements/recoveries and whole-block operations. The final operation is a swap accomplished via a single measurement/recovery¹⁸.

computer is $(5n + 4)(3 + K/k)$. The ratio

$$S = \frac{5n + 4}{k} (1 + 3k/K) \quad (2)$$

is the scale-up in computer size necessary to allow fault tolerance by this method.

I need to estimate the number of elementary gates and time steps required to prepare an ancillary n -bit block in the state $|0\rangle_L$, and to verify the state so that the only X errors which remain in it are uncorrelated with each other. I assume that operations on separate blocks can be carried out in parallel, and so can single-bit operations within a block. One further type of parallelism will be allowed, namely that a multiple controlled-NOT, involving one control bit and several target bits in the same block, can be performed in a single time step. This is physically reasonable as it is possible in implementations such as the ion trap, in which a communal degree of freedom is coupled to every qubit in a block, and transitions in many qubits can be driven simultaneously. This assumption was not used previously¹⁰. It allows the number of time steps required to build $|0\rangle_L$ to equal to the number of rows in $\bar{H}$, which is $(n - k)/2$. Also, as parity checks are measurements of $\bar{X}_u$ and $\bar{Z}_u$ operators, they

Table 1 Code performance comparison

Code	P (10^{-17})	γ (10^{-6})	ϵ (10^{-6})	$(5n + 4)/k$
$[[127, 29, 15]]$	48	10	0.15	22
$[[255, 143, 15]]$	240	6	0.05	9
$[[127, 43, 13]]$	71	5	0.08	15
$[[99, 5, 15]]$	8	10	0.20	100
$[[47, 1, 11]]$	1.7	3	0.14	239
$[[79, 1, 15]]$	1.7	10	0.25	399

Error rates and scale-up required to run a quantum algorithm of size $KQ = 7.5 \times 10^{15}$. The first column gives the parameters of the quantum code; the second gives the required success probability for recovery of a single block (equation (6)); the third and fourth give the required error rates (equations (3)–(5)), and the final column gives the scale-up in computer size (equation (2)). The first three codes are BCH, the last three are quadratic residue.

can be performed by controlled multiple-NOTS from the verification bit onto the ancilla. All the parity checks in $\tilde{H}$ and $\tilde{D}$ are evaluated, making $(n+k)/2$ in all. Allowing two time steps per check, I need $(n+k)$ to verify one ancilla. A further $4n$ steps are used for the bitwise CX and CZ between ancillas and data block.

One full recovery of the computer fails if any block develops more than $t \equiv (d-1)/2$ errors, or a wrong syndrome is adopted. The former is dominant when the syndrome is extracted $O(t+1)$ or more times¹⁰. For blocks whose first extracted syndrome was non-zero, the probability to develop more than t errors is¹⁰

$$P \approx 2 \sum_{i=t+1}^g \frac{g!}{i!(g-i)!} \left(\frac{2}{3} \gamma + \frac{s}{g} \frac{2}{3} \epsilon \right)^i \quad (3)$$

where γ is the probability of gate failure, ϵ is the probability of memory error per qubit per time step, and g and s are the number of independent opportunities for gate and memory errors, respectively. The errors either occur directly in the block to be corrected, or they originate in an ancilla and are propagated into the block. I find¹⁰

$$g \approx n(4r+1) \quad (4)$$

$$s \approx n((3n+k)/2 + 2 + n(4+r/2)) \quad (5)$$

where r is the number of repetitions of the syndrome extraction required on average for confidence that one has the right syndrome. I will take $r = t+1$, which is a safe overestimate¹⁰.

The only remaining possibility is that a zero syndrome is wrongly obtained for some block during the first round of syndrome extraction, and then further noise generates an uncorrectable error before the next recovery. Such a block accumulates noise for twice the normal time, and is coupled to $t+2$ ancillas; the probability for $t+1$ errors is thus $< 2^{t+1}P$. To wrongly obtain a zero syndrome requires an error in the syndrome which matches the syndrome, an event with probability small compared to $1/2^{t+1}$, so this failure rate is negligible.

To realise Q Toffoli gates on a K -qubit quantum computer, it is necessary that

$$P < \frac{k}{8KQ} \quad (6)$$

as $8Q$ successful recoveries of each block are required.

Table 1 shows the scale-up and error rates needed to satisfy equations (3) to (6) for various QECCs, for $KQ = 7.5 \times 10^{15}$. This size of computation is sufficient to factorize a 1,000-digit (3,300-bit) number using Shor's algorithm¹¹. I find that the $[[127, 29, 15]]$ code permits such a task with scale-up 22 and $\gamma = 10^{-5}$, $\epsilon = 1.5 \times 10^{-7}$.

Fault-tolerant quantum computing can work well with efficient quantum error correcting codes such as the $[[127, 29, 15]]$ CSS code obtained from the classical $[127, 78, 15]$ BCH code. The success relies on the ability to merge useful measurements on the logical state with recovery operations, on careful network design and on optimized use of ancillas. The resulting physical computer is only an order of magnitude larger than the logical machine contained within it. To take advantage of these results, a quantum computing system would need to be amenable to repeated measurement and parallel operation of quantum gates. $\square$

Received 27 October 1998; accepted 5 March 1999.

1. Steane, A. M. Quantum computing. *Rep. Prog. Phys.* **61**, 117–173 (1998).
2. Steane, A. M. Introduction to quantum error correction. *Phil. Trans. R. Soc. Lond. A* **356**, 1739–1758 (1998).
3. Shor, P. W. Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A* **52**, R2493–R2496 (1995).
4. Steane, A. M. Error correcting codes in quantum theory. *Phys. Rev. Lett.* **77**, 793–797 (1996).
5. Calderbank, A. R. & Shor, P. W. Good quantum error-correcting codes exist. *Phys. Rev. A* **54**, 1098–1105 (1996).
6. Steane, A. M. Multiple particle interference and quantum error correction. *Proc. R. Soc. Lond. A* **452**,

2551–2577 (1996).

7. Knill, E. & Laflamme, R. A theory of quantum error correcting codes. *Phys. Rev. A* **55**, 900–911 (1997).
8. Bennett, C. H., DiVincenzo, D. P., Smolin, J. A. & Wootters, W. K. Mixed state entanglement and quantum error correction. *Phys. Rev. A* **54**, 3822–3851 (1996).
9. Knill, E. & Laflamme, R. Concatenated quantum codes. Preprint quant-ph/9608012 at (<http://www.lanl.gov>) (1996).
10. Steane, A. M. Space, time, parallelism and noise requirements for reliable quantum computing. *Fortschr. Phys.* **46**, 443–457 (1998).
11. Preskill, J. Reliable quantum computers. *Proc. R. Soc. Lond. A* **454**, 385–410 (1998).
12. Knill, E., Laflamme, R. & Zurek, W. H. Resilient quantum computation: error models and thresholds. *Proc. R. Soc. Lond. A* **454**, 365–384 (1998).
13. Shor, P. W. in *Proc. 37th Symp. on Foundations of Computer Science* 15–65 (IEEE Computer Soc. Press, Los Alamitos, CA, 1996).
14. Kitaev, A. Yu. Quantum computations: algorithms and error correction. *Russ. Math. Surveys* **52**, 1191–1249 (1997).
15. DiVincenzo, D. P. & Shor, P. W. Fault-tolerant error correction with efficient quantum codes. *Phys. Rev. Lett.* **77**, 3260–3263 (1996).
16. Steane, A. M. Active stabilisation, quantum computation and quantum state synthesis. *Phys. Rev. Lett.* **78**, 2252–2255 (1997).
17. Knill, E. Group representations, error bases and quantum codes. Preprint quant-ph/9608049 at (<http://xxx.lanl.gov>) (1996).
18. Gottesman, D. A theory of fault-tolerant quantum computation. *Phys. Rev. A* **57**, 127–137 (1998).
19. Gottesman, D. Class of quantum error correcting codes saturating the quantum Hamming bound. *Phys. Rev. A* **54**, 1862–1868 (1996).
20. Calderbank, A. R., Rains, E. M., Shor, P. W. & Sloane, N. J. A. Quantum error correction and orthogonal geometry. *Phys. Rev. Lett.* **78**, 405–409 (1997).
21. Steane, A. M. Enlargement of Calderbank Shor Steane quantum codes. *IEEE Trans. Inf. Theory* (in the press); preprint quant-ph/9802061 at (<http://xxx.lanl.gov>) (1998).
22. Gottesman, D. Fault-tolerant quantum computation with higher-dimensional systems. Preprint quant-ph/9802007 at (<http://xxx.lanl.gov>) (1998).
23. Knill, E. Personal communication.
24. Beckman, D., Chari, A. N., Devabhaktuni, S. & Preskill, J. Efficient networks for quantum factoring. *Phys. Rev. A* **54**, 1034–1063 (1996).
25. Vedral, V., Barenco, A. & Ekert, A. Quantum networks for elementary arithmetic operations. *Phys. Rev. A* **54**, 147–153 (1996).
26. Steane, A. M. Efficient fault-tolerant quantum computing. Preprint quant-ph/9809054 at (<http://xxx.lanl.gov>) (1998).
27. MacWilliams, F. J. & Sloane, N. J. A. *The Theory of Error-correcting Codes* (North-Holland, Amsterdam, 1996).
28. Grassl, M., Beth, Th. & Pellizzari, T. Codes for the quantum erasure channel. *Phys. Rev. A* **56**, 33–38 (1997).
29. Zalka, C. Threshold estimate for fault tolerant quantum computing. Preprint quant-ph/9612028 at (<http://xxx.lanl.gov>) (1996).

Acknowledgements. The author is supported by The Royal Society and St Edmund Hall, Oxford.

Correspondence and requests for materials should be addressed to the author (e-mail: a.steane@physics.ox.ac.uk).

Three-dimensional orientation measurements of symmetric single chromophores using polarization microscopy

S. A. Empedocles, R. Neuhauser & M. G. Bawendi

Department of Chemistry and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

A complete understanding of any complex molecular system generally requires a knowledge of the three-dimensional (3D) orientation of its components relative both to each other, and to directional perturbations such as interfaces and electromagnetic fields. Far-field polarization microscopy is a convenient and widespread technique for detecting and measuring the orientation of single chromophores. But because the polarized electromagnetic field that is used to probe the system lacks a significant longitudinal component, it was thought that, in general, only 2D orientation information could be obtained^{1–3}. Here we demonstrate that far-field polarization microscopy can yield the 3D orientation of certain highly symmetric single chromophores (CdSe nanocrystal quantum dots in the present case). The key requirement is that the chromophores must have a degenerate transition dipole oriented isotropically in two dimensions, which gives rise to a perpendicular 'dark axis' that does not couple to the

light field. By measuring the fluorescence intensity from the dipole as a function of polarization angle, it is possible to calculate both the tilt angle between the dark axis and the sample plane, as well as the in-plane orientation, and hence obtain the 3D orientation of the chromophore

The 3D orientation of individual transition dipoles can be determined in excitation using near-field scanning optical microscopy (NSOM), as a result of the spatially varying longitudinal and transverse polarized electromagnetic field emanating from the near-field tip⁴. Although it is a powerful application of surface scanning probe microscopy, NSOM is generally more invasive, more complex, and less routinely applicable than far-field microscopy. In addition, NSOM only collects data from one region of the sample at a time, precluding the highly parallel data acquisition possible with far-field imaging methods⁵.

Although far-field microscopy is appealing for many single chromophore applications, and polarization spectroscopy can be a very precise measure of orientation³, polarized electromagnetic waves in the far-field have a strong electromagnetic field in only one direction: such waves can, in general, only provide orientation information about the projection of a transition dipole within the sample plane (the plane normal to the propagation direction of the light)¹⁻³. This was thought to be a limitation of far-field polarization microscopy in applications where 3D orientation is of interest. Non-polarization techniques can measure chromophores oriented out of the sample plane, taking advantage of lifetime changes at a dielectric interface², optical aberrations⁶ or special microscope objectives⁷. These techniques, however, generally rely on the precise modelling of specific extrinsic experimental factors rather than the intrinsic polarization properties of the chromophore.

The strength of an optical transition is proportional to $|\boldsymbol{\mu} \cdot \mathbf{E}|^2$, where $\boldsymbol{\mu}$ is the transition dipole vector and $\mathbf{E}$ is the polarization of

the absorbed or emitted light. Transition dipoles in the single chromophores studied so far are generally unidirectional, creating a single 'bright axis' which couples to the absorbed or emitted electromagnetic field. The strength of such a transition is proportional to $\cos^2(\theta)\cos^2(\phi)$, where θ is the angle between the polarization of the light and the projection of $\boldsymbol{\mu}$ onto the sample plane, and ϕ is the tilt angle between $\boldsymbol{\mu}$ and the sample plane. If a polarizer is rotated in the emission pathway, the intensity of the collected fluorescence oscillates between $I_{\min} = 0$ and $I_{\max} = |\boldsymbol{\mu}|^2 \cdot \cos^2(\phi)$ (the same is true in excitation as well). The result is 100% polarized absorption and emission, regardless of the orientation of the dipole. Although $I_{\max}$ is a qualitative measure⁸ of ϕ , local inhomogeneities make it difficult to know $|\boldsymbol{\mu}|$, so that the out-of-plane angle cannot generally be determined. As a result, for systems with a 'bright axis', standard far-field polarization spectroscopy can typically only provide information about the orientation of the projection of a single transition dipole within the sample plane (in-plane orientation).

Highly symmetric molecules can have a degenerate transition dipole oriented isotropically in a plane (for instance, benzene has a degenerate transition dipole in the plane of the aromatic ring⁹ ($S_00^0 \leftarrow S_16^1$)). These systems are better characterized by a unique unidirectional 'dark axis', that is oriented normal to the transition dipole plane, and that does not couple to the light field. Other examples of systems with a potential 'dark axis' include the family of highly symmetric triphenyl methane dyes (for example, crystal violet) which have been used as biological tags and have also been studied on the single-molecule level¹⁰. The 2D nature of a degenerate transition dipole can be exploited to provide information about the 3D orientation of 'dark axis' chromophores using standard far-field optics.

The strength of a 'dark axis' transition is proportional to

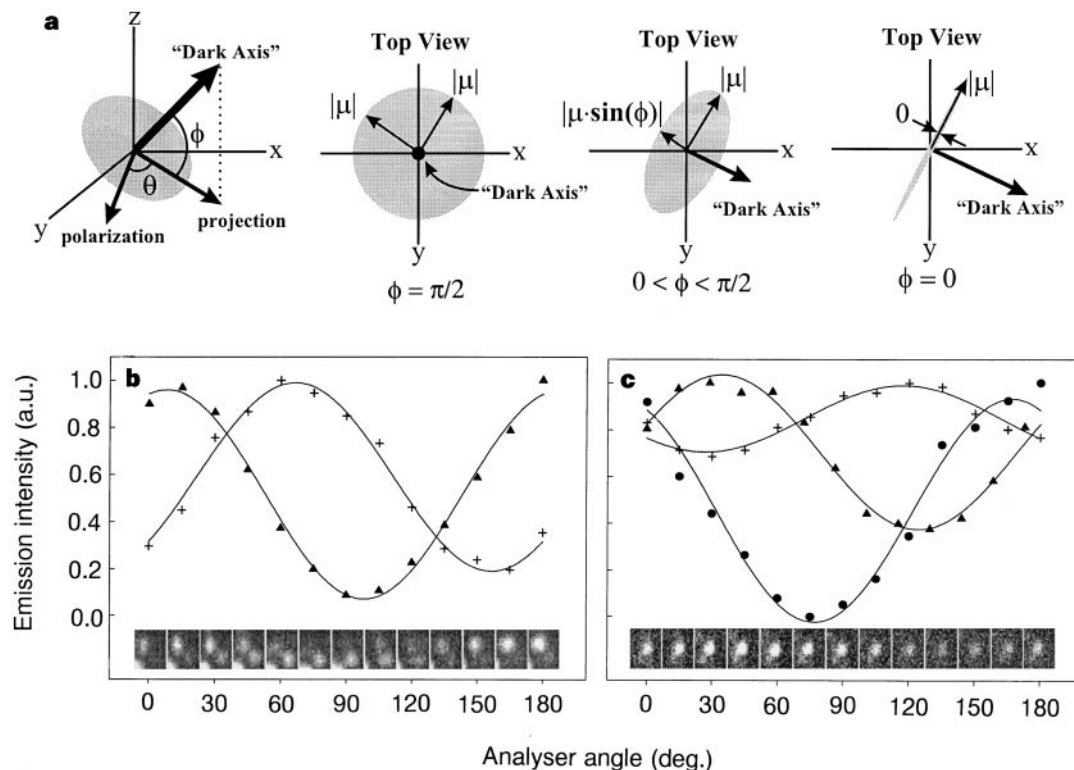


Figure 1 Polarization dependence. **a**, 'Dark axis' coordinate system with light propagation along the z-axis. The shaded disk represents the planar, degenerate transition dipole oriented normal to the 'dark axis'. Top views depict the projection of the dipole onto the sample (x-y) plane at three different values of ϕ , corresponding to the polarization distribution observed by the detection system. Also included are the magnitudes of two orthogonal dipole projections,

as well as the projection of the 'dark axis' onto the sample plane at each angle. **b**, Normalized emission intensity of two single NCs as a function of polarization angle with fit. Inset, images of the NCs in **b** as a function of angle, which can be read directly off of the x-axis. **c**, Polarization data for three additional NCs with the corresponding images for one (triangles). Plotted data in **b** and **c** are the average of 2 and 6 complete 180° rotations, respectively.

$(1 - \cos^2(\theta)\cos^2(\phi))$ where θ and ϕ are now defined relative to the 'dark axis' orientation (Fig. 1a). In this case, rotation of the emission polarization causes the strength of the transition to oscillate between $I_{\max} = |\mu|^2$ and $I_{\min} = |\mu|^2 \sin^2(\phi)$. Since $I_{\max}$ is measured directly, the value of ϕ can, in principle, be calculated (we note, however, that there is no distinction between angles into (ϕ) and out of ($-\phi$) the sample plane). As a result, it should be possible to use far-field polarization spectroscopy to obtain information about the 3D orientation of single 'dark axis' chromophores. This result is not specific to a particular chromophore, but should be true for any system with a degenerate transition dipole.

One system that has been extensively studied on the single chromophore level and is predicted to have a 'dark axis' is nanocrystals (NCs) of CdSe^{5,11,12}. These nanometre-sized crystals represent a chemical route to the formation of semiconductor quantum dots, and have attracted growing interest in recent years due to their strongly size-dependent optical properties that can be tuned during synthesis¹³. The slightly prolate shape (aspect ratio 1.1–1.2) and the unidirectional nature of the wurtzite crystal structure of these NCs results in an electronic structure that is predicted to have many 'dark axis' transition dipoles^{14–16}.

Effective mass approximation¹⁵ and pseudopotential¹⁷ calculations predict that emission from the lowest excited state in CdSe NCs is a spin-forbidden transition. As a result, relaxation from this state is thought to proceed through either a virtual transition or

mixing with a nearby optically active state¹⁴. Tight-binding approximation calculations also predict mixing of the lowest excited state with higher-lying optically active states¹⁸. Of the three optically active states near the band edge, two potential transition dipole orientations exist¹⁵. One is a 'bright axis' dipole oriented along the *c*-axis of the NC. The other is a degenerate dipole oriented isotropically in the *x*–*y* plane, resulting in a *c*-oriented 'dark axis'. Transitions involving a 'bright axis' dipole should have a polarization dependence similar to what has been observed in previous single-molecule experiments, with a high degree of polarization regardless of the NC orientation. In contrast, transitions involving a 'dark axis' dipole should have a wide range of polarizations depending on the orientation of the 'dark axis'.

The inset of Fig. 1b shows images of two single CdSe NCs oscillating out of phase with each other as a function of polarization angle. Also plotted are the normalized emission intensities as a function of angle. The intensities oscillate with a high degree of polarization $[(I_{\max} - I_{\min})/I_{\max}]$ as a sine-squared function of angle. The high degree of polarization confirms the existence of a strongly oriented transition dipole. Rotation of the excitation polarization had no observable effect on the polarization of the emitted light, indicating that the observed polarization in emission is not polarization memory, but is instead consistent with an intrinsic, oriented transition dipole. While Fig. 1b demonstrates the highly oriented nature of the transition dipole, further examination reveals that the degree of polarization varies strongly between different NCs (Fig. 1c). This is consistent with emission from a 'dark axis' dipole, with a distribution of out-of-plane orientations.

Figure 2 displays data from 176 single NCs, showing the phase and degree of polarization for each. A broad distribution of parameters is observed which covers the entire range of phase angles and polarizations. To confirm that the differences observed in Figs 1 and 2 are not artefacts of the experimental or sample preparation procedures (such as those described in ref. 1), single DiI molecules were also studied. As expected for a system with a 'bright axis'², the average degree of polarization for the 10 single DiI molecules studied was found to be ~100% (Fig. 2 inset). Results identical to those in Figs 1 and 2 have also been obtained for NCs deposited in hexane, precluding a contribution from the surrounding matrix¹.

Whereas emission is characterized by a broad range of polarizations, the polarization dependence of the excitation was found to be significantly weaker (average 23%, $\sigma = 14\%$). Although strong absorption polarization has been reported for single CdS NCs¹², the

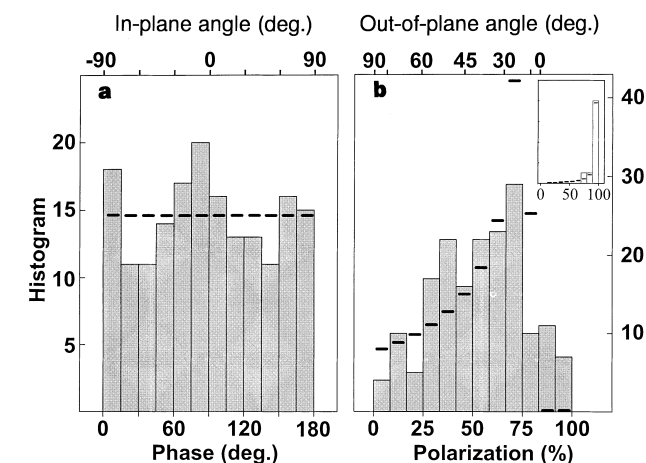


Figure 2 Polarization statistics. **a**, Histogram of polarization phase angles for 176 single NCs, with the corresponding in-plane orientation of the NC *c*-axis (top axis). **b**, Histogram of degrees of polarization with the corresponding out-of-plane angle for the same 176 NCs. Inset, histogram of degrees of polarization of 10 single DiI molecules. Solid lines in the main and inset figure correspond to the calculated probability histogram for an isotropic distribution of 'dark' and 'bright' axis orientations, respectively. The data here are consistent with emission from a 'dark axis' dipole with an isotropic distribution of orientations within the ensemble. Inset data (DiI) is representative of a distribution of 'bright axis' dipoles. To relate polarization to orientation, two additional effects must be considered. First, the radiation pattern of a dipole on a surface differs from a dipole in free space²⁴. Second, the large collection angle of the microscope (NA = 0.7) allows the detection of some emitted radiation polarized parallel to the collection axis. It is therefore possible to find 'bright axis' transitions with less than 100% polarization. The degree of polarization for 'dark axis' transitions is also reduced relative to the simple model. The effects of the surface and collection angle are included in the calculated probability histogram as well as the top axes of this figure. An additional consequence of the large collection angle is that the degree of polarization for 'bright axis' transitions should be strongly correlated with $I_{\max}$ since dipoles with the lowest degrees of polarization also emit the fewest photons toward the detector. Consistent with emission from a 'dark axis' dipole, where $I_{\max}$ should have negligible angular dependence, examination of 85 single NCs revealed no such correlation. (DiI is 1,1'-diocetadecyl-3,3',3'-tetramethylindocarbocyanine.)

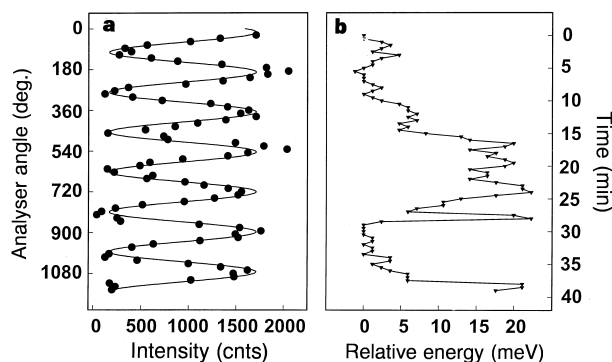


Figure 3 Polarization versus spectral diffusion. **a**, Total emission intensity of a single NC as a function of polarization angle with fit. **b**, Relative emission energy of the zero phonon line of the NC in **a** as a function of time. The data in **a** and **b** were taken simultaneously. Over 40 min, while many large spectral diffusion shifts can be seen, no change in the degree of polarization or phase is observed. Since the transition dipole is not affected by electric fields, the data here, in conjunction with ref. 11, strongly suggest that spectral diffusion is the result of a dynamic local electric field and not reorientation of the NC within a static field.

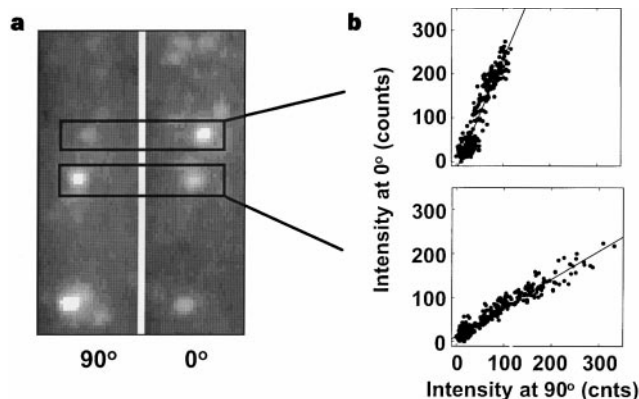


Figure 4 Polarized emission at room temperature. **a**, Two images of the same single NCs taken by projecting the 0° (right) and 90° (left) components of the emission polarization simultaneously onto different areas of the CCD detector. NCs which appear brighter on the right side have a transition dipole oriented more strongly along the 0° axis. The opposite is true for NCs oriented along the 90° axis. **b**, Scatter plots for two NCs where the intensity of emission polarized along the 0° axis is plotted against the intensity along the 90° axis. Although large fluctuations in the total emitted intensity are observed over time at room temperature, the ratio of the intensities remains the same. The slope of each plot is a stable measure of the relative polarization along each axis. Data here were taken with a 0.5-s integration time.

weak excitation polarization observed in the present experiment is not unexpected, given that the NCs were excited in a region with a relatively high density of electronic states. While individual states should have distinct transition dipole orientations^{14,15}, it is likely that multiple, overlapping states are excited simultaneously, decreasing the degree of polarization in excitation.

There may be additional contributions to the observed emission polarization dependence, such as the prolate shape of the NCs. In porous silicon, differential screening of the electromagnetic field along the major and minor axes of a dielectric ellipsoid has been shown to produce stronger emission and absorption along the major axis¹⁹. On its own, however, this effect can only account for less than 20% polarization in CdSe NCs and therefore cannot explain the large degree of polarization observed. In addition, since the *c*-axis is oriented parallel to the major axis of the NC, differential screening due to shape should not contribute to the polarization dependence of either 'bright' or 'dark' axis transitions, because emission from these states occurs either along the major or minor axes, but not both. Shape may play a more significant role in excitation polarization where the strong electronic orientation effects are reduced.

Another possible contribution to the observed polarization dependence is large local electric fields ($\sim 10^5$ V cm⁻¹) that previous Stark measurements have suggested exist around single NCs¹¹. To determine the role of local fields in the orientation of the transition dipole, the polarization dependence of single NCs was studied in the presence and absence of a large electric field (2.5×10^5 V cm⁻¹) applied within the sample plane¹¹. These experiments found that in all single NCs studied, the phase and degree of polarization was not changed by the applied field, suggesting that local electric fields do not contribute to the orientation of the transition dipole.

The data in Figs 1 and 2 suggest that at 10 K, band edge emission in CdSe NCs proceeds through a 'dark axis' transition, allowing us to estimate both the in-plane and out-of-plane angles for each NC. It is therefore possible to use the highly parallel and flexible detection of far-field polarization microscopy to directly measure the 3D orientation of each NC within the sample. The top axes of Figs 2a and b have been added to indicate the corresponding angles.

Knowledge of orientation relative to the experimental frame of reference can be valuable in interpreting single NC experiments. We

demonstrate this by studying spectral diffusion. Spectral diffusion (spectral shifting over time) is a common characteristic of many single chromophore systems including CdSe NCs^{5,11,12}. In NCs, the mechanism is thought to involve changing local electric fields, which produce Stark shifts in the spectra of single NCs¹¹. In the experiments of ref. 11, however, there was no way to distinguish between a changing local field and changes in the orientation of a NC relative to a static local field. This distinction is important since it addresses the nature of the dynamics in this system (that is, is the local field changing relative to the NC or is the NC rotating relative to the local field?). By simultaneously monitoring polarization and spectral diffusion (Fig. 3), we find that the NCs do not rotate during spectral shifts, suggesting that it is the local field that changes over time.

Many theoretical predictions that can be investigated on the single NC level rely on knowing the orientation of the *c*-axis relative to applied perturbations such as pressure, and electric and magnetic fields. Studies of coupling or energy transfer between NCs, or between a NC and a surface, may also benefit from the knowledge of relative orientations. Although the experiments described here were performed at cryogenic temperatures, measurements of single 38-Å NCs at room temperature also revealed strong emission polarization (Fig. 4). Recent work has demonstrated that NCs can be used to label biological systems^{20,21}. These results suggest the possibility of using single NCs at chemically and biologically relevant temperatures to monitor orientation and conformation changes in systems such as proteins and polymers. □

Methods

Polarization dependence measurements. NCs with an average diameter of 52 Å were synthesized as described in ref. 13 and were then overcoated with a ~ 10 -Å layer of ZnS to increase their fluorescence quantum yield^{22,23}. Single NC samples were prepared as in ref. 5. Images and spectra were taken at 10 K using a far field epifluorescence microscope, also described in ref. 5. Data were taken by rotating a polarizer/depolarizer pair in front of the detection system. The optical system was tested in both image and spectral mode by confirming that no polarization dependence was observed in randomly oriented ensemble NC samples as well as single 1-μm dye spheres (Nile red), under circular polarized excitation. Excitation polarization was controlled by passing circularly polarized light through a rotatable linear polarizer. Polarization at the sample surface was greater than 200:1. Excitation for all data was with 60 W cm⁻² of 514-nm light. All images and spectra (not shown) were taken with 5- and 30-s integration times, respectively.

Received 21 October 1998; accepted 26 February 1999.

- Guttlar, F. *et al.* Single molecule polarization spectroscopy: pentacene in p-terphenyl. *Chem. Phys.* **211**, 421–430 (1996).
- Macklin, J. J. *et al.* Imaging and time-resolved spectroscopy of single molecules at an interface. *Science* **272**, 255–258 (1996).
- Ha, T. *et al.* Single molecule dynamics studied by polarization modulation. *Phys. Rev. Lett.* **77**, 3979–3982 (1996).
- Betzig, E. & Chichester, R. Single molecules observed by near-field scanning optical microscopy. *Science* **262**, 1422–1425 (1993).
- Empedocles, S. A., Norris, D. J. & Bawendi, M. G. Photoluminescence spectroscopy of single CdSe nanocrystallite quantum dots. *Phys. Rev. Lett.* **77**, 3873–3876 (1996).
- Dickson, R. M., Norris, D. J. & Moerner, W. E. Simultaneous imaging of individual molecules aligned both parallel and perpendicular to the optic axis. *Phys. Rev. Lett.* **81**, 5322–5325 (1998).
- Sepiol, J., Jasny, J., Keller, J. & Wild, U. P. Single molecules observed by immersion mirror objective. The orientation of terrylene molecules via the direction of its transition dipole moment. *Chem. Phys. Lett.* **273**, 444–448 (1997).
- Bopp, M. A. *et al.* Single-molecule spectroscopy with 27 fs pulses: time-resolved experiments and direct imaging of orientational distributions. *Appl. Phys. Lett.* **73**, 7–9 (1998).
- Callomon, J. H., Dunn, T. M. & Mills, I. M. Rotational analysis of the 2600 Å absorption system of benzene. *Phil. Trans. R. Soc. Lond. A* **259**, 499–532 (1966).
- Kneipp, K. *et al.* Single molecule detection using surface-enhanced raman scattering (SERS). *Phys. Rev. Lett.* **78**, 1667–1670 (1997).
- Empedocles, S. A. & Bawendi, M. G. Quantum-confined Stark effect in single CdSe nanocrystallite quantum dots. *Science* **278**, 2114–2117 (1997).
- Tittel, J. *et al.* Investigations of the emission properties of single CdS-nanocrystallites. *Ber. Bunsenges. Phys. Chem.* **101**, 1626–1630 (1997).
- Murray, C. B., Norris, D. J. & Bawendi, M. G. Synthesis and characterization of nearly monodisperse CdE (E = S, Se, Te) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706–8715 (1993).
- Efros, A. L. Luminescence polarization of CdSe microcrystals. *Phys. Rev. B* **46**, 7448–7458 (1992).
- Efros, A. L. *et al.* Band-edge exciton in quantum dots of semiconductors with a degenerate valence band: dark and bright exciton states. *Phys. Rev. B* **54**, 1–14 (1996).
- Shaing, J. J. *et al.* Symmetry of annealed wurtzite CdSe nanocrystals: assignment to the C_{3v} point group. *J. Phys. Chem.* **99**, 17417–17422 (1995).

17. Franceschetti, A., Fu, H., Wang, L. W. & Zunger, A. Many-body pseudopotential theory of excitons in InP and CdSe QDs. *Phys. Rev. B* (in the press).
18. Leung, K., Pokrant, S. & Whaley, K. B. Exciton fine structure in CdSe nanoclusters. *Phys. Rev. B* **57**, 12291–12301 (1998).
19. Kovalev, E. *et al.* Optically induced polarization anisotropy in porous Si. *Phys. Rev. Lett.* **77**, 2089–2092 (1996).
20. Bruchez, M. Jr *et al.* Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**, 2013–2016 (1998).
21. Chan, C. W. & Nie, S. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* **281**, 2016–2018 (1998).
22. Hines, M. A. & Guyot Sionnest, P. Synthesis and characterization of strongly luminescent ZnS-capped CdSe nanocrystals. *J. Phys. Chem.* **100**, 468–471 (1996).
23. Dabbousi, B. O. *et al.* (CdSe)/ZnS core-shell quantum dots: synthesis and characterization of a size series of highly luminescent nanocrystallites. *J. Phys. Chem. B* **101**, 9463–9475 (1997).
24. Fetting, C. & Lukosz, W. Optical-environment-dependent lifetimes and radiation patterns of luminescent centers in very thin films. *J. Lumin.* **31&32**, 933–935 (1984).

Acknowledgements. We thank A. P. Efron for discussions. S.A.E. thanks the Lester Wolfe Foundation and Eastman Chemical Co. for fellowships. This work was funded in part by the NSF-MRSEC program and the AT&T Foundation. We thank the M.I.T. Harrison Spectroscopy laboratory for support and for use of its facilities.

Correspondence and requests for materials should be addressed to M.G.B. (e-mail: mgb@mit.edu).

Size and form in efficient transportation networks

Jayanth R. Banavar*, Amos Maritan† & Andrea Rinaldo‡

* Department of Physics and Center for Materials Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

† International School for Advanced Studies (SISSA), Via Beirut 2–4, 34014 Trieste, and INFN and the Abdus Salam International Center for Theoretical Physics, 34014 Trieste, Italy

‡ Ralph M. Parsons Laboratory, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA, and Dipartimento di Ingegneria Idraulica, Marittima e Geotecnica, Università di Padova, Padova, Italy

Many biological processes, from cellular metabolism to population dynamics, are characterized by allometric scaling (power-law) relationships between size and rate^{1–10}. An outstanding question is whether typical allometric scaling relationships—the power-law dependence of a biological rate on body mass—can be understood by considering the general features of branching networks serving a particular volume. Distributed networks in nature stem from the need for effective connectivity¹¹, and occur both in biological systems such as cardiovascular and respiratory networks^{1–8} and plant vascular and root systems^{1,9,10}, and in inanimate systems such as the drainage network of river basins¹². Here we derive a general relationship between size and flow rates in arbitrary networks with local connectivity. Our theory accounts in a general way for the quarter-power allometric scaling of living organisms^{1–10}, recently derived⁸ under specific assumptions for particular network geometries. It also predicts scaling relations applicable to all efficient transportation networks, which we verify from observational data on the river drainage basins. Allometric scaling is therefore shown to originate from the general features of networks irrespective of dynamical or geometric assumptions.

In euclidean geometry, a D -dimensional compact object, characterized by a linear size L and having a constant density (independent of L), has a volume V and a mass M that scale as L^D . Thus, simple geometrical attributes that depend on L should scale with the mass as a function of $M^{1/D}$. For example, the surface area of such objects scales as $M^{(D-1)/D}$. For three-dimensional objects (where $D = 3$), one may therefore expect scaling to hold with the exponents being related to the factor $1/3$. But we will show here that, for systems comprising transportation networks, this simple $1/D$ -scaling is no longer valid.

To model the metabolic system of living organisms, we postulate that the fundamental processes of nutrient transfer at the micro-

scopic level are independent of organism size. In a D -dimensional organism (denoted the service region), the number of such transfer sites scales as L^D (here L is measured in units of l , the mean distance between neighbouring sites). Each transfer site is fed with nutrients (for example, through blood) by a central source through a network providing a route for the transport of the nutrients to the sites. The total amount of nutrients being delivered to the sites per unit time, B , simply scales as the number of sites or as L^D . The total blood volume C for a given organism at any given time depends, in the steady-state supply situation, on the structure of the transportation network. It is proportional to the sum of individual flow rates in the links or bonds that constitute the network. We define the most efficient class of networks as that for which C is as small as possible. Note that this does not coincide with the assumption made in ref. 8, where the energy dissipation was minimized within a hierarchical model.

Our key result is that, for networks in this efficient class, C scales as $L^{D/(D+1)}$. The total blood volume increases faster than the metabolic rate B as the characteristic size scale of the organism increases. Thus larger organisms have a lower number of transfer sites (and hence B) per unit blood volume. Because the organism mass scales^{1–3,5–7} (at least) as C , the metabolic rate does not scale linearly with mass, but rather scales as $M^{D/(D+1)}$. In the non-biological context, the number of transfer sites is proportional to the volume of the service region, which, in turn, leads to a novel mass–volume relationship.

We consider a single network source that services L^D sites uniformly distributed in a D -dimensional space. Each site is connected to one or more of its neighbours, which results in a transportation network that spans the system. Such a network may be a well connected one with loops, or merely a spanning tree, an extreme example of which is a spiral structure¹¹ (Fig. 1a–d). Each site X is supplied by the source at a steady rate F_X , no less than a positive value $F_{\min}$ and no larger than a value $F_{\max}$ ($F_{\max} \geq F_{\min}$), both of which may depend on l . A simple, special case would correspond to a uniform constant rate for all sites. We set $l = 1$ without loss of generality. Such a system could represent a biological organism that needs a steady supply of nutrients to all its parts^{1–8}. The metabolic rate of such a biological organism is given by $B = \sum_X F_X$ and simply scales as L^D . Another example is the (inverse) problem of the drainage basin of a river where the sites represent source areas, F_X is the net rainfall rate of landscape-forming events¹² and the network provides routes for transport of water and sediments.

Any transportation network must provide a route from the source to all the L^D sites and consists of interconnected links in each of which, in steady state, the flow rate does not change with time. Each link starts or terminates at the source or a site. Let the scalar quantity $|I_b|$ represent the magnitude of the flow on the b -th link. The source has an outward flow, whose rate exactly equals the sum of the flow rates into all the sites. At a junction of the links, a conservation law for the net flow holds—the inflow must exactly balance the outflow plus the amount supplied to the site. This conservation law does not determine uniquely the flow on each link for an arbitrary network. The degrees of freedom in the choice of the flow pattern is controlled by the number of independent loops equal to (the number of links) – (the number of sites) + 1. The total quantity of nutrients in the network at any instant of time, C , is given simply by $\sum_b |I_b|$ on setting the constant of proportionality equal to 1.

We propose the theorem that for any spanning network in D dimensions, C scales at least as L^{D+1} and at most as L^{2D} for large L . Equivalently, the number of transfer sites or B scales at most as $C^{D/(D+1)}$ and at least as $C^{1/2}$. The general proof of the theorem is presented as Supplementary Information. A simple argument is that C is equal to the number of transfer sites (L^D) multiplied by the mean distance of the transfer sites to the source (the father away a transfer site is from the source as measured along the network, the larger the amount of blood that will be needed to nourish it). This

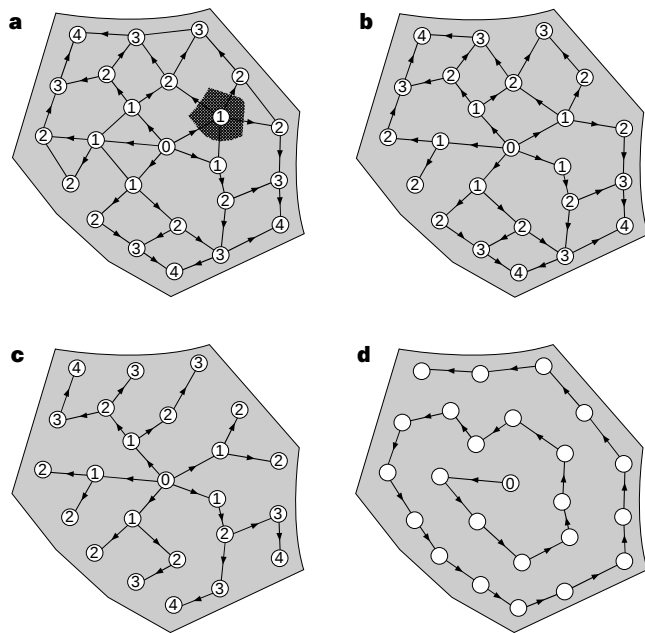


Figure 1 Sketches of transportation networks. **a**, An example of a fully connected network, defined as the connection pattern of links that join the sites, generically denoted by X , characterized by different distances L_X from the source O . The node number in the drawing represents the value of L_X defined to be the minimum number of sites encountered among all the routes along the network from O to X . The cross-hatched area represents the elementary service volume V_X of one of the sites. Each site X may be thought of as serving an elementary volume V_X requiring the necessary nutrient supply F_X . Two sites are defined as neighbours if their service volumes share a part of their boundaries of non-zero measure. The arrows denote orientated links that are directed away from the source. **b**, Maximal directed network. Only orientated links are retained from the fully connected network. **c**, A directed spanning tree. Each site is connected to the source by a single path, the shortest, belonging to the maximal network. **d**, A spiral pattern, which yields limiting scaling behaviour. Note that within the assumed framework of local connectivity, the possibility of explosion patterns¹¹ connecting all sites directly to the source is not admissible.

mean distance scales at least as L and at most as L^D . The latter result holds for a network with a one-dimensional topology as in a space-filling spiral, whereas the former result is obtained when the network has all links directed away from the source (or towards a collection point or the outlet in the river basin). Indeed, C scales as L^{D+1} for all directed networks (the flow in all but the links of the directed maximal network (Fig. 1b) are necessarily zero in these cases) independent of whether they have loops or a tree-like structure, provided I_b is non-negative on each link. Such solutions do indeed exist and include all directed trees (Fig. 1c). These solutions belong to the class of the most efficient networks in that they lead to the smallest value of C . In such networks, B scales as $C^{D/(D+1)}$. Thus, for $D = 3$, when quantities related to the network are scaled with respect to C , we obtain quarter-power scaling rather than the one-third power behaviour discussed earlier.

The application of our theorem to the problem of allometric scaling in living organisms, which span size scales that range over many orders of magnitude^{1–8}, is straightforward. In spite of an impressive array of scales and the accompanying diverse requirements in the resources needed for sustaining the organism, a robust and common feature is that a variety of biological quantities (generically denoted as Y) that are related to blood circulation scale algebraically with the mass M of the organism. The relation is given as $Y \sim M^s$, where s is a scaling exponent^{1–8}. Even though the organisms are three-dimensional, the exponent s is usually consistently found to be obtained from the fraction $1/4$.

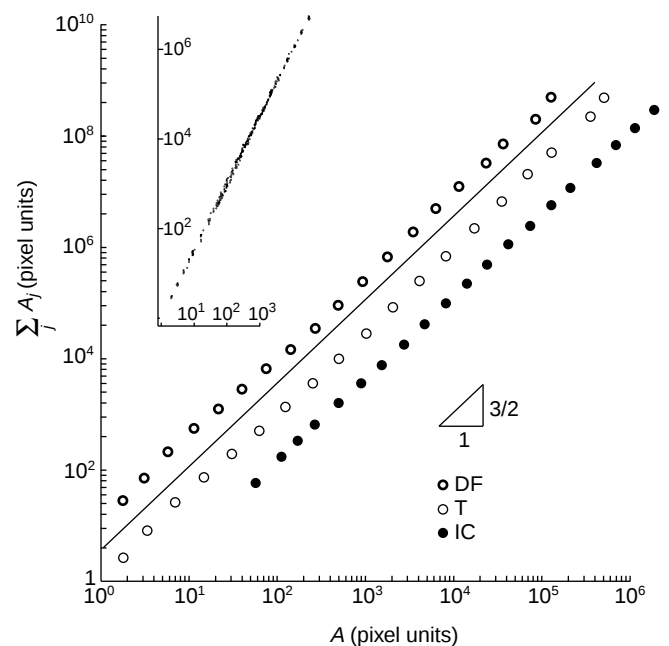


Figure 2 Allometric scaling in river networks. Double logarithmic plot of $C = \Sigma A_i / A_X$ versus A for three river networks characterized by different climates, geology and geographic locations (Dry Fork, West Virginia, 586 km², digital terrain map (DTM) size 30×30 m²; Island Creek, Idaho, 260 km², DTM size 30×30 m²; Tirso, Italy, 2,024 km², DTM size 237×237 m²). The experimental points are obtained by binning total contributing areas, and computing the ensemble average of the sum of the inner areas for each sub-basin within the binned interval. The figure uses pixel units in which the smallest area element is assigned a unit value. Also plotted is the predicted scaling relationship with slope $3/2$. The inset shows the raw data from the Tirso basin before any binning has been done.

West, Brown and Enquist⁸ have constructed a model of space-filling hierarchical networks of branching tubes to explain allometric scaling. In our analysis, the mass M of an organism scales as the blood volume C , so that in the simplest and most efficient scenario $B \sim M^{3/4}$, which is the central result of allometric scaling^{1–10}. Many of the other exponents derived in ref. 8 follow from simple dimensional analysis, thus accounting for their robustness, whereas others depend on detailed assumptions. The scaling exponent is universal. Our analysis shows that the basic result does not require any assumptions regarding the hierarchical nature of the network nor does it necessarily demand a tree-like structure. However, the presence of a tree would greatly shorten the total length of the network, thereby increasing its viability and efficiency. Observed differences in scaling within a species and between species^{1–7} could arise from factors extrinsic to the network that limit the amount of nutrients delivered to the sites (see Supplementary Information).

We now turn to a test of the theorem within the context of river networks¹². An elevation map of the soil heights of the rugged landscape may be used to derive a spanning tree that defines unique routes from each location within the basin to the global outlet, where the main stream is formed. Suitably accurate data and objective procedures to extract the network are known, and the reliability of the observational results are well established¹². Each site X in the basin is associated with a sub-basin that drains into it. The role of the metabolic rate for this sub-basin is taken by the total contributing area, which is defined by the recursion relation $A_X = \Sigma_{Z \in nm(X)} A_Z + 1$, where $nm(X)$ are the nearest neighbours of X that drain into X through appropriate steepest-descent drainage directions. Note that the added unity is the area of the elementary pixel, the analogue of F_X . Indeed, if A is the area of the sub-basin that drains into a given site X , the analogue of C is defined by $\Sigma_{Z \in \gamma} A_Z$

where γ is the collection of all sites connected to X through drainage directions. The theorem predicts that, for efficient drainage basins, a log-log plot of C versus A should have a slope of $3/2$ because $D = 2$ and river networks in nature are known to be efficient and directed¹². Our observational data (Fig. 2) are found to agree with the predictions over five decades of scales.

Our general results should be applicable to a wide variety of distributed networks including the flow of water, blood, sewage, food, air and electrical currents. Even though the specific details vary significantly, the novel behaviour built into efficient transportation networks provides a unified framework¹³ underlying the allometric scaling of diverse systems. □

Received 17 August 1998; accepted 10 March 1999.

1. McMahon, T. A. & Bonner, J. T. *On Size and Life* (Scientific American Library, New York, 1983).
2. Bonner, J. T. *The Evolution of Complexity by Means of Natural Selection* (Princeton Univ. Press, Princeton, 1983).
3. Peters, R. H. *The Ecological Implications of Body Size* (Cambridge Univ. Press, Cambridge, 1983).
4. Feldman, H. A. & McMahon, T. A. The $3/4$ mass exponent for energy metabolism is not an artifact. *Respir. Physiol.* **52**, 149–163 (1983).
5. Schmidt, Nielsen, K. *Scaling: Why is Animal Size so Important?* (Cambridge Univ. Press, Cambridge, 1984).
6. Calder, W. A. III *Size, Function and Life History* (Harvard Univ. Press, Cambridge, Massachusetts, 1984).
7. Brown, J. H. *Macroecology* (Univ. Chicago Press, Chicago, 1995).
8. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
9. Enquist, B. J., Brown, J. H. & West, G. B. Allometric scaling of plant energetics and population density. *Nature* **395**, 163–165 (1998).
10. Damuth, J. D. Common rules for animals and plants. *Nature* **395**, 115–116 (1998).
11. Stevens, P. S. *Patterns in Nature* (Little, Brown, Boston, 1974).
12. Rodriguez-Iturbe, I. & Rinaldo, A. *Fractal River Basins: Chance and Self-Organization* (Cambridge Univ. Press, New York, 1997).
13. Wilson, E. O. *Consilience: The Unity of Knowledge* (Knopf, New York, 1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank R. Rigon for the computations shown in Fig. 2; A. Beauvais, F. Colaiori, P. Dodds, A. Flammini, M. Caterina Putti, R. Robinett, I. Rodriguez-Iturbe, D. Rothman and J. Weitz for helpful discussions; and John Damuth for many key insights. This work was supported by INFN, NASA, NATO, MURST 40% *Trasporto di sedimenti ed evoluzione morfologica di corsi d'acqua, estuari e lagune alle diverse scale temporali* and The Donors of the Petroleum Research Fund administered by the American Chemical Society.

Correspondence and requests for materials should be addressed to J.R.B. (e-mail: jayanth@phys.psu.edu).

Electronic mechanism of hardness enhancement in transition-metal carbonitrides

Seung-Hoon Jhi*, Jisoon Ihm*, Steven G. Louie† & Marvin L. Cohen†

* Department of Physics and Center for Theoretical Physics, Seoul National University, Seoul 151-742, Korea

† Department of Physics, University of California at Berkeley, Berkeley, California 94720–7300, USA and Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

Transition-metal carbides and nitrides are hard materials widely used for cutting tools and wear-resistant coatings. Their hardness is not yet understood at a fundamental level. A clue may lie in the puzzling fact that transition-metal carbonitrides that have the rock-salt structure (such as $\text{TiC}_x\text{N}_{1-x}$) have the greatest hardness for a valence-electron concentration of about 8.4 per cell^{1–3}, which suggests that the hardness may be determined more by the nature of the bonding than by the conventional microstructural features that determine the hardness of structural metals and alloys. To investigate this possibility, we have evaluated the shear modulus of various transition-metal carbides and nitrides using *ab initio* pseudopotential calculations. Our results show that the behaviour of these materials can be understood on a fundamental level in terms of their electronic band structure. The unusual hardness originates from a particular band of σ bonding states between the

non-metal p orbitals and the metal d orbitals that strongly resists shearing strain or shape change. Filling of these states is completed at a valence-electron concentration of about 8.4, and any additional electrons would go into a higher band which is unstable against shear deformations.

Among the various microscopic and intrinsic properties of materials, the shear modulus provides a measure of the rigidity against the shape deformations involved in microhardness indentation experiments. In particular, since the Peierls stresses in the transition-metal carbonitrides are very high, the strengths of these compounds may be influenced more by the difficulty of nucleating and moving dislocations through the background crystal lattice than by the difficulty of moving dislocations through microstructural obstacles⁴. As the stresses required to nucleate or move isolated dislocations scale with the shear modulus, electronic changes that affect the shear modulus may have a pronounced effect on the macroscopic hardness value. Several prior discussions of the mechanical properties of hard materials have made this point^{4–8}. There is more than one shear modulus, but we have elected to study c_{44} (rather than, say, the difference between c_{11} and c_{12} or averaged shear modulus) which by itself represents a shape change without volume change, and provides directly information about electronic response to shear strain.

To understand the variation of the shear modulus c_{44} with composition, we carried out quantum-mechanical electronic-structure calculations (under both normal and strained conditions) for $\text{TiC}_x\text{N}_{1-x}$, $\text{HfC}_x\text{N}_{1-x}$ and $\text{Zr}_x\text{Nb}_{1-x}\text{C}$ using the *ab initio* pseudopotential method^{9,10}. In our calculations, alloy configurations such as $\text{TiC}_x\text{N}_{1-x}$ are simulated in two different ways and cross-checked. The first is the virtual crystal method in which the ionic pseudopotential

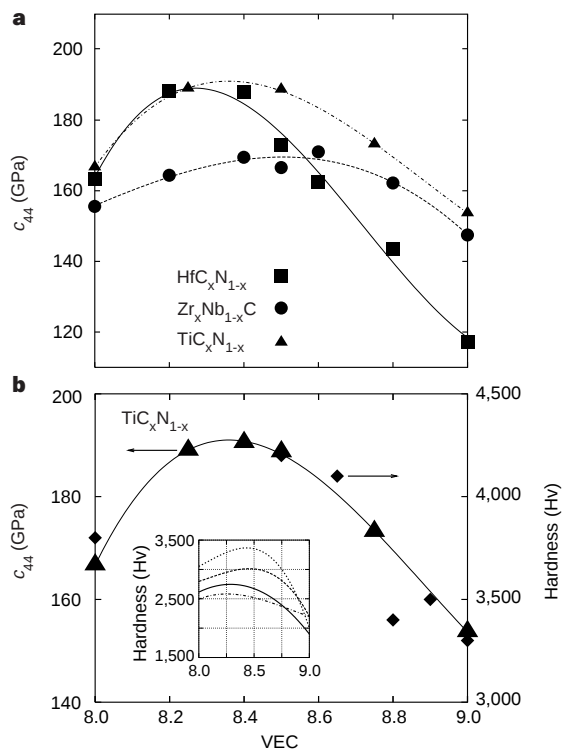


Figure 1 Correlation between the calculated shear modulus c_{44} and experimental microhardness. **a**, Calculated shear modulus c_{44} in GPa for $\text{TiC}_x\text{N}_{1-x}$ (filled triangles), $\text{HfC}_x\text{N}_{1-x}$ (filled boxes) and $\text{Zr}_x\text{Nb}_{1-x}\text{C}$ (filled circles) as a function of the valence electron concentration (VEC). The curves are polynomial fits to the calculated c_{44} . **b**, Measured microhardness of $\text{TiC}_x\text{N}_{1-x}$ cermet from ref. 2 in Hv units (filled diamonds). For comparison, the calculated c_{44} (filled triangles) in GPa is plotted in the same figure. The thin solid line is a guide to the eye. Inset, microhardness of bulk alloys and sub-stoichiometric compounds (Ti(CN) , dotted line; ZrNbC , dashed line; Hf(CN) , solid line; and NbC_{1-x} , dash-dotted line).

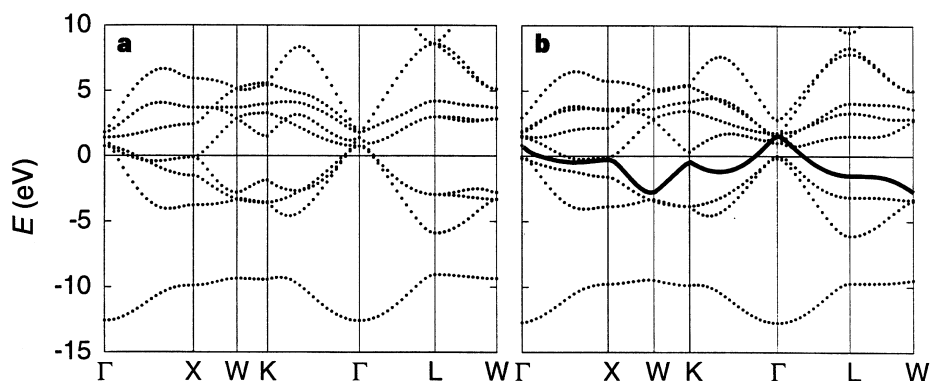


Figure 2 Influence of shear strain on the electronic structure of TiC. **a**, Electronic band structure at zero strain. **b**, Electronic band structure at finite shear strain ($\epsilon_{xy} = 0.1$). For comparison, the band structure of sheared TiC is plotted along the symmetry line equivalent to that of the face-centred cubic (f.c.c.) structure. The Fermi level lies at the energy zero. If carbon atoms are replaced by nitrogen atoms in $\text{TiC}_x\text{N}_{1-x}$ alloys, extra electrons from nitrogen will occupy the empty states just

above the Fermi level at first, and then fill higher-lying metallic conduction bands as the nitrogen concentration increases further. The fourth band shows a larger rise in energy on the application of the shear stress (solid line in **b**) compared with its unstrained counterpart in **a**. The fifth band in **b**, by contrast, is lowered in energy compared with that in **a**.

for, say, C_xN_{1-x} is approximated by the weighted average of those of C and N and then the usual *ab initio* self-consistent pseudopotential calculation is done. The second is a supercell method, with a set of particular atomic arrangements chosen to mimic random alloys. No assumption is made about the electronic structure, and a fully self-consistent calculation is performed. The two methods yield results within a few per cent of each other and the trend found is identical.

Figure 1a shows the calculated shear modulus as a function of the valence-electron concentration (VEC). For all cases shown ($\text{TiC}_x\text{N}_{1-x}$, $\text{Zr}_x\text{Nb}_{1-x}\text{C}$ and $\text{HfC}_x\text{N}_{1-x}$), $x = 1$ corresponds to a VEC of 8 (electrons per unit cell) and $x = 0$ to a VEC of 9. The shear modulus c_{44} shows a maximum value for VEC in the range 8.3–8.5. In Fig. 1b, we show together the experimental microhardness (in Vickers hardness scale, Hv) and the calculated c_{44} of $\text{TiC}_x\text{N}_{1-x}$ to demonstrate the close correlation between the two. Experimental hardnesses for other materials are plotted in the inset to show that the trends are essentially the same. To understand in more detail the behaviour of c_{44} common to all these materials, we choose $\text{TiC}_x\text{N}_{1-x}$ as a representative material and discuss the change in its electronic structure under shear strain.

Figure 2 shows the band structure of TiC at equilibrium and under the shear strain of $\epsilon_{xy} = \epsilon_{yx} = 0.1$. Three states derived from carbon $2p$ orbitals, degenerate at the Γ -point, split into two inequivalent states of $p_x + p_y$ and p_z symmetry under the shear strain. The metal d orbitals of e_g symmetry also split into $d_{3z^2-r^2}$ and $d_{x^2-y^2}$ states in the second and fourth band, respectively, of the strained electronic band structure. As seen in the figure, drastic change occurs along the symmetry line $\text{K}-\Gamma-\text{L}$ near the Fermi level: we note the conspicuous rise of the fourth band relative to other bands, under the shear strain. The movement of the fourth band is closely related to its highly directional bonding character (Fig. 3a). In the region near K, the fourth band is mainly derived from σ bonding interactions between the titanium $d_{x^2-y^2}$ orbitals and the carbon p_x and p_y orbitals. Optimal bonding occurs when the Ti–C–Ti bond angle is 90° . Under the shear strain, this angle deviates from the optimal geometry and the energy level of the fourth band rises to second order in ϵ_{xy} . The large increase in the electronic energy under shear strain shows that electrons occupying the fourth band are extremely resistive to shear. The second and third bands mainly consist of π bonding between p and d orbitals which is relatively insensitive to the shear distortion. But the metallic σ bonding states in the fifth band, derived from $d-d$ interactions, are very sensitive to the shear strain. Due to the strengthening of the metal–metal bonding along the $[1\bar{1}0]$ direction by the shortening of the inter-metallic distance under shear strain (Fig. 3b), this band is lowered

giving rise to a negative contribution to the elastic shear modulus if it is occupied. Along the $\Gamma-\text{L}$ line, a similar analysis holds, but the bonding character is complicated by the influence of the third-nearest-neighbour atoms.

The strain energy density for a symmetric shear strain $e_{xy} = \epsilon_{xy} + \epsilon_{yx}$ is given by:

$$u(e_{xy}) = \frac{1}{2} c_{44} e_{xy}^2 \quad (1)$$

As the energy rise of the fourth band is second order in the shear strain ϵ_{xy} , electrons occupying this band contribute positively to the shear modulus. As x decreases from 1 in the $\text{TiC}_x\text{N}_{1-x}$ alloys, additional electrons provided by the nitrogen atoms occupy the empty states of the fourth band near the Γ -point (degenerate with the second and third bands in the absence of the strain) and enhance c_{44} . At $x \approx 0.6$, corresponding to VEC = 8.4, the Fermi level lies at the local valence band maximum at the Γ -point and the fourth band is completely filled. When the carbon concentration (expressed as x)

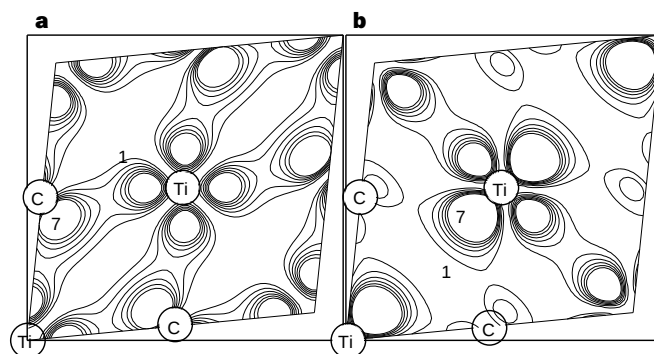


Figure 3 Charge density of a state of TiC near the K-point of the Brillouin zone under shear strain ($\epsilon_{xy} = 0.1$). **a**, Charge-density plot of the fourth band on the (001) plane. **b**, Charge-density plot of the fifth band on the same plane as in **a**. When shear strain is absent, the bonding character of the fourth band is covalent $pd\sigma$ bonding between titanium (Ti) and carbon (C), and that of the fifth band is weak $dd\sigma$ metallic bonding. The contour line step is 1 electron per unit cell, and the charge density increases from zero (in the interstitial region) to 7 electrons per unit cell (around atomic sites). As the shear strain distorts the bond axis of the $pd\sigma$ bonding in the fourth band and raises the energy level, electrons occupying the fourth band give a positive contribution to the elastic shear modulus. On the other hand, electrons in the fifth band give a negative contribution to the shear modulus because the shear strain enhances the bond strength of metallic $dd\sigma$ bonding and lowers the energy level.

decreases well below 0.6, a local gap opens at the Γ -point and the band structure looks similar to that of TiN. The decrease of c_{44} at $x \leq 0.6$ is attributed in large part to the filling of the fifth band near the K-point which gives a large negative contribution to c_{44} as mentioned above. Hence the occurrence of the shear-modulus maximum at the VEC of ~ 8.4 originates from the complete filling of the shear-resistive $p\delta\sigma$ bonding states at this electron concentration. The occurrence of the experimental microhardness maximum at the same VEC strongly suggests the same origin for the hardness (see below).

Generally speaking, hardness is determined by the mobility of dislocations. A central question is then whether the creation and motion of dislocations in our materials can be correlated with a simpler and more fundamental property, such as the shear modulus. For the covalently bonded materials, the very low mobility of dislocation kinks is the rate-determining factor for dislocation motion and is in turn determined by the very high bond-breaking energy under a large shear strain. The bond-breaking energy involved in plastic deformation and the bond-restoring energy under elastic shear strain essentially go together in covalently bonded hard materials. This reasoning indicates that the response of well-localized bonds (such as covalent bonds) to shear is the crucial factor in determining hardness. For many $s-p$ tetrahedrally bonded semiconductors, their hardness also shows a strong correlation with the bulk modulus, which has been a focus of some previous theoretical work¹¹. The shear modulus, however, encompasses a greater set of materials as a general indicator of hardness.

The microhardness of the alloys considered here is about 17–23 GPa, approximately one-eighth of the calculated c_{44} . In indentation tests, materials deform typically with about 8–12% strain, on average², and the hardness is of the same magnitude as the yield stress. The ratio of the hardness to the calculated c_{44} indicates that the yield stress or the critical shear stress of the alloys is close to the theoretical value of the perfect crystals ($1/2\pi$ times the shear modulus), which is a common feature of covalently bonded materials¹³. The highly directional coupling between metal d and non-metal $2p$ electrons results in a shear-resistive covalent bonding in the present case. The kink migration energy is typically a few tenths of the cohesive energy for covalently bonded materials (for bulk silicon, it is¹⁴ ~ 1.6 eV, about one-third of the experimental cohesive energy of 4.63 eV), exceeding the typical strain energy in indentation tests. Under such circumstances, dislocations are hard to activate at low temperature, and the critical shear stress is expected to be close to that of the perfect crystal. In short, in the plastic deformation of these materials, slip occurs in a manner expected of a perfect crystal, and the shear modulus is a good measure to quantify the resistance to plastic deformation.

Using this argument, other transition-metal alloys with the rock-salt structure are expected to have maximum hardness at the same VEC (~ 8.4). In fact, it has been experimentally verified¹ that $\text{Ti}_x\text{Nb}_{1-x}\text{C}$ has a hardness trend similar to that of $\text{Zr}_x\text{Nb}_{1-x}\text{C}$. Moreover, our theory may be used to explain the hardness behaviour of the sub-stoichiometric group V carbides. The hardness of such carbides is found^{1,3} to have a maximal value at a specific vacancy concentration of $\sim 12\%$. If the electronic structures of sub-stoichiometric compounds are not much changed from those of the stoichiometric phase, the maximal hardness is expected at the vacancy concentration which gives a VEC of 8.4; that is, 12.5%, in agreement with the above observation. However, one should be cautious about generalizing this argument, as a significant deviation from stoichiometry may alter the shape and character of the band structure. In addition, the behaviour of the hardness (and correspondingly, the elastic shear modulus) of sub-stoichiometric transition-metal compounds may depend on the microstructure of the vacancies; for example, aggregates of vacancies¹⁵ may occur. Some sub-stoichiometric nitrides indeed show different hardness trends with varying vacancy concentrations^{1,16}.

We note that the bulk modulus of metallic hard materials does not show a maximum between VEC values of 8 and 9. The anomalous maximum in hardness or c_{44} of these materials is attributed to the presence near the Fermi level of two electronic bands that respond oppositely to shear stress. Under hydrostatic pressure, on the other hand, symmetry is preserved and all bands are resistive to volume change. The bulk modulus has been shown theoretically¹⁷ to be a monotonic function of the VEC. The presence of bands of two different bonding characters near the Fermi level makes the hard transition-metal alloys distinct from semiconductors or insulators where the bulk modulus, shear modulus, and hardness all show the same trends. $\square$

Received 11 December 1998; accepted 22 March 1999.

- Hollock, M. Material selection for hard coatings. *J. Vac. Sci. Technol. A* **4**, 2661–2669 (1986).
- Richter, V., Beger, A., Drobniowski, J., Endler, I. & Wolf, E. Characterization and wear behavior of TiN- and $\text{TiC}_x\text{N}_{1-x}$ -coated cermets. *Mater. Sci. Eng. A* **209**, 353–357 (1996).
- Toth, L. E. *Transition Metal Carbides and Nitrides* (Academic, New York, 1971).
- Kelly, A. & Macmillan, N. H. *Strong Solids* 3rd edn (Clarendon, Oxford, 1986).
- Gilman, J. J. The plastic resistance of crystals. *Aust. J. Phys.* **13**, 327–346 (1960).
- Clerc, D. G. & Ledbetter, H. M. Mechanical hardness: a semiempirical theory based on screened electrostatics and elastic shear. *J. Phys. Chem. Solids* **59**, 1071–1095 (1998).
- Teter, D. M. Computational alchemy: the search for new superhard materials. *Mater. Res. Soc. Bull.* **23**, 22–27 (1998).
- Krenn, C. R., Morris, J. W., Jhi, S.-H. & Ihm, J. In *Hard Coatings Based on Borides, Carbides & Nitrides* (eds Kumar, A., Chung, Y.-W. & Chia, R. W. J.) 379–388 (The Minerals, Metals & Materials Soc., Warrendale, PA, 1998).
- Ihm, J., Zunger, A. & Cohen, M. L. Momentum-space formalism for the total energy of solids. *J. Phys. C* **12**, 4409–4422 (1979).
- Pickett, W. E. Pseudopotential methods in condensed matter applications. *Comput. Phys. Rep.* **9**, 115–197 (1989).
- Liu, A. Y. & Cohen, M. L. Prediction on new low compressibility solids. *Science* **245**, 841–842 (1989).
- Samuels, L. E. in *Microindentation Technique in Materials Science and Engineering* (eds Blau, P. J. & Lawn, B. R.) 5–24 (ASTM, Philadelphia, 1985).
- Gilman, J. J. Physical chemistry of intrinsic hardness. *Mater. Sci. Eng. A* **209**, 74–81 (1996).
- Nikitenko, V., Farber, B. & Lunin, Yu. Experimental investigation of the dynamics of kinks of dislocation lines in semiconductor single crystals. *Sov. Phys. JETP* **66**, 738–746 (1987).
- Hull, D. & Bacon, D. J. *Introduction to Dislocations* 3rd edn (Pergamon, Oxford, 1984).
- Jiang, X. et al. Elastic constants and hardness of ion-beam-sputtered TiN_x films measured by Brillouin scattering and depth-sensing indentation. *J. Appl. Phys.* **69**, 3053–3057 (1991).
- Jhi, S.-H. & Ihm, J. Electronic structure and structural stability of $\text{TiC}_x\text{N}_{1-x}$ alloys. *Phys. Rev. B* **56**, 13826–13829 (1997).

Acknowledgements. S.-H.J. and J.I. were supported by the SNI, the SRC programme of KOSEF, and the BSRI programme of KRF. S.G.L. and M.L.C. were supported by the NSF and DOE. The support of a Korea-US Cooperative Grant by KOSEF and NSF is acknowledged. S.G.L. acknowledges the hospitality of the Korea Institute for Advanced Study, where part of this Letter was written.

Correspondence and requests for materials should be addressed to S.G.L. (e-mail: louie@uclink4.berkeley.edu).

Near-field probing of vibrational absorption for chemical microscopy

B. Knoll & F. Keilmann

Max-Planck-Institut für Biochemie, D-82152 Martinsried, Germany

Identification of chemical compounds by vibrational spectroscopy at infrared wavelengths requires macroscopic samples: the spatial resolution is diffraction-limited to a scale of about half the wavelength, or about five micrometres. The scanning near-field optical microscope^{1,2}, however, can reveal sub-wavelength detail because it uses near-field probing rather than beam focusing. Here we demonstrate the use of the aperture-less approach to scanning near-field optical microscopy^{3–6} to obtain contrast in vibrational absorption on a scale of about 100 nanometres, about one-hundredth of a wavelength. We record infrared scattering from the tip of an atomic force microscope scanned over a composite polymer film. At the boundary between different polymers we observe contrast changes owing to changes in vibrational absorption. The contrast is strongly enhanced in the near field of the probe tip, which we interpret as evidence of surface-enhanced infrared absorption⁷. When extended to multi-wavelength opera-

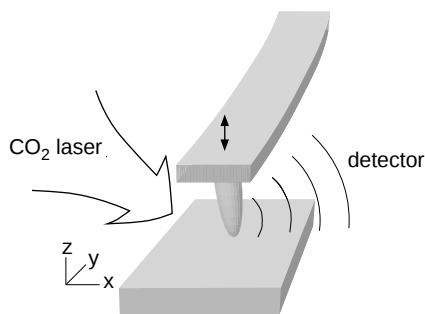


Figure 1 Aperture-less mid-infrared microscope. A dithering metallized AFM tip scatters infrared radiation, with λ selectable between 9 and 11 μm , onto a detector while a sample is being scanned in close proximity.

tion, this approach should enable imaging of chemical composition at nanometre resolution.

Infrared spectroscopists do not routinely obtain micrometre-scale detail because diffraction limits the resolution to a few micrometres. Because the improved technique of the aperture-scanning near-field optical microscope^{8,9} also suffers from a wavelength-related limit (at about $\lambda/10$, because of the difficulty of squeezing light through a small aperture), its infrared realization¹⁰ has no potential to attain sub-micrometre resolution.

The radically different aperture-less design of the near-field probe has enabled much better wavelength-relative resolution below $\lambda/100$ —and in one case, even $\lambda/10^6$ (ref. 11)—to be attained. In this approach, the aperture is replaced by a much smaller particle or antenna tip which serves as a field-concentrating scattering centre³. Obviously the (free-space) illumination of such a scatterer has at least a wavelength-sized cross-sectional diameter, obeying Abbe's rule, and generates a background scattering from all over the tip structure. This signal is, however, stationary. In contrast, scattering from the tip apex depends sensitively on the presence of a nearby sample; modulating the tip-to-sample separation d (ref. 4) therefore allows one to electronically filter the tip's near-field response from the background and extract the high-resolution content.

Our experiment (Fig. 1) employs commercial Si cantilevered tips (NT-MDT, Moscow, Russia) which we metallize with a 100-nm layer of gold to obtain a quasi-electrostatic infrared field concentration at the tip apex. The cantilever is 'dithered' perpendicularly to the sample surface at its resonance frequency (40 kHz) by a piezoelectric transducer; a free amplitude of 120 nm peak-to-peak is measured by a standard optical deflection system. On approaching the sample, the amplitude decreases and is stabilized at 80% by electronic feedback (tapping atomic force microscope, AFM, mode). AFM and infrared images are simultaneously recorded. The infrared beam from a tunable CO₂ laser (MPB, Québec, Canada) is attenuated to 10 mW and polarized by non-deviating metal-grid elements (Lasnix, Berg, Denmark), and focused by a Cassegrain NA = 0.55 objective (Ealing Electro-Optics, Watford, UK) onto the tip. Fine adjustment of the focusing is critical to obtain contrast in the infrared image. A concave mirror collects near-forward scattered radiation onto a mercury cadmium telluride detector (Infrared Associates, Stuart, USA). This signal after lock-in amplification constitutes the near-field response. Figure 2 shows the result of imaging a well-defined metallic test sample; this sample was prepared by evaporating a 15-nm layer of gold onto a silicon wafer partly covered by adsorbed polystyrene (PS) spheres which were subsequently removed.

We model the infrared scattering by considering the electrical dipole moment of the tip (regarded as a sphere to simplify the calculations) together with its mirror image in the sample. Our near-field model does not take account of the metal cone supporting the sphere, as we assume the cone has no influence on relative signal magnitudes, that is, on the infrared contrast itself; but the cone

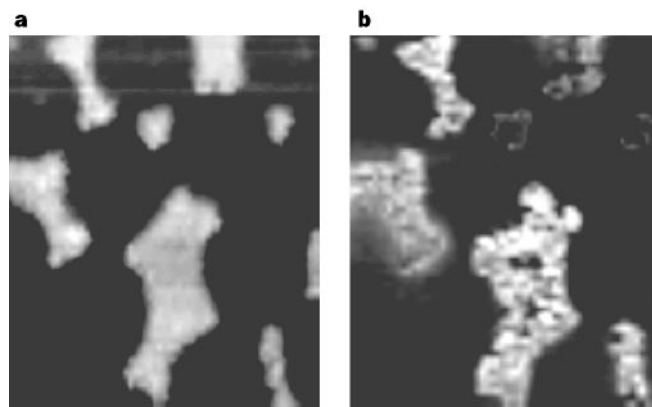


Figure 2 Images of gold islands on silicon. The islands are 15 nm high, and the field of view is 900 × 1100 nm. **a**, AFM topography; **b**, simultaneous infrared image at $\lambda = 10.6 \mu\text{m}$

certainly increases absolute signals through far-field antenna effects of coupling and polarizing, imposing for example a longitudinal electric field direction to the tip (ref. 11). The near-field interaction of a sphere of radius a —fitting the curvature of the apex—with dielectric constant ϵ_t (tip) at a distance d from a halfspace characterized by the dielectric constant ϵ_s (sample) is governed by the effective polarizability of

$$\alpha_{\perp}^{\text{eff}} = \frac{\alpha(1 + \beta)}{1 - \frac{\alpha\beta}{16\pi(a+d)^3}}, \quad (1)$$

$$\text{where } \beta = \frac{\epsilon_s - 1}{\epsilon_s + 1} \quad \text{and} \quad \alpha = 4\pi a^3 \frac{\epsilon_t - 1}{\epsilon_t + 2}$$

for the applied longitudinal polarization. The resulting extinction cross-section arises from two contributions, one $C_{\text{abs}} = k\text{Im}\{\alpha_{\perp}^{\text{eff}}\}$ (where Im stands for imaginary part) representing non-radiative loss ($k = 2\pi/\lambda$), the other $C_{\text{sca}} = k^4|\alpha_{\perp}^{\text{eff}}|^2/6\pi$ radiative scattering¹². In our experiment we use sharp tips with $a \ll 100$ nm and thus obtain $C_{\text{sca}} \ll C_{\text{abs}}$, so C_{sca} can be neglected when measuring near-forward scattering. For $a = 30$ nm and $\lambda = 10 \mu\text{m}$ we find $C_{\text{abs}} = 0.03 \text{ nm}^2$ assuming $\epsilon_s = 1$, that is, for Mie scattering of an isolated Au sphere. In close proximity ($d < a$) to Au or intrinsic Si, the cross-section C_{abs} is augmented by a factor 5 or 3, respectively,

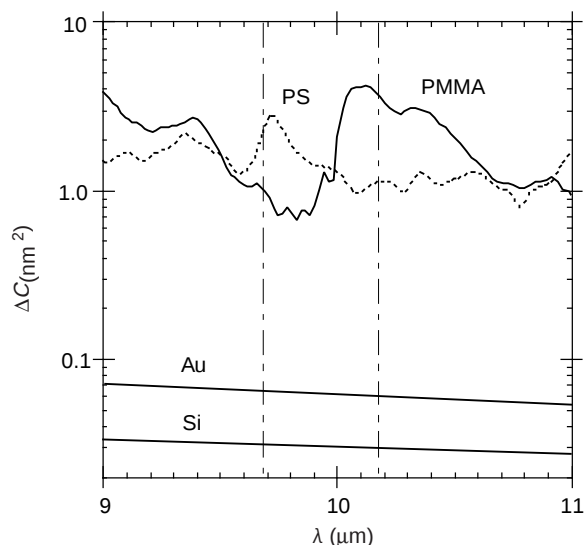


Figure 3 Near-field extinction cross-section ΔC . The cross-section is calculated for an Au sphere of $2a = 60$ nm diameter close to a dielectric or metallic material.

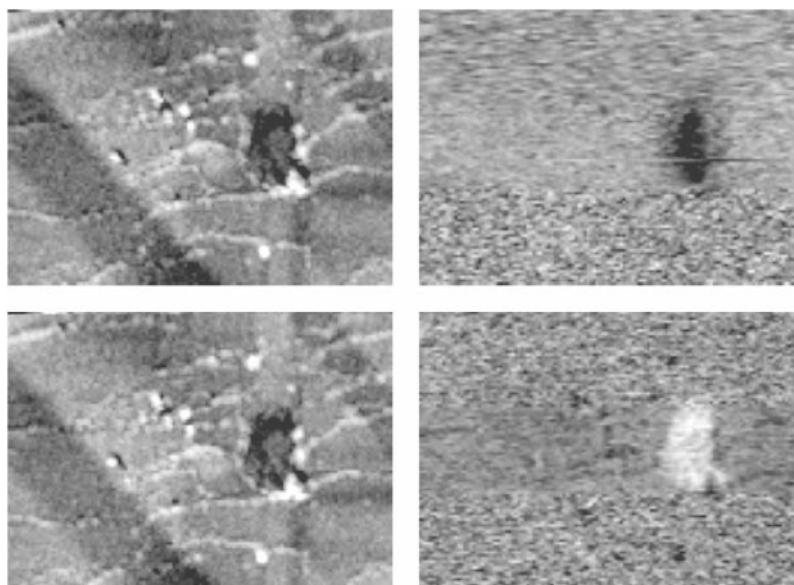


Figure 4 Infrared images (right) of PS embedded in PMMA, and simultaneously recorded AFM topography (left). The upper right image is at a wavelength

$\lambda = 9.68 \mu\text{m}$, and the lower right at $10.17 \mu\text{m}$, demonstrating wavelength-dependent vibrational contrast. The field of view is $3.5 \times 2.5 \mu\text{m}$.

and by as much as 400 for vibrationally active PS or polymethylmethacrylate (PMMA) polymers. The effect drops with increasing separation d to about 60% at $d = 120 \text{ nm}$, and then remains nearly constant. We calculate the cross-sectional difference defined as $\Delta C = C_{\text{abs}}(d = 0) - C_{\text{abs}}(d = 120 \text{ nm})$, which represents the observable near-field response (Fig. 3).

Our experiment directly records ΔC on a somewhat larger background of modulated scattering ΔC_b from the metal cone beyond the tip, the phase of which depends on fine adjustment of the infrared illumination. Although this prevents absolute values and signs of contrast from being determined, we find the contrast between Au and Si of 20–30% (Fig. 2) to be smaller than that calculated (Fig. 3), as expected. The Au response (Fig. 2) is very sensitive to the tip's position within the unstructured islands. The infrared image probably reveals a hidden crystallite substructure¹³. The sharp metal boundaries imaged in Fig. 2 give evidence of the high resolving power ($<30 \text{ nm}$) of the infrared microscope. The substructure in the absence of a topographic counterpart proves that topography-related artefacts are not significant.

To assess the infrared contrast offered by vibrational resonances, we image a sample composed of two immiscible polymers, PS and PMMA. The complex dielectric function of the pure materials was measured by ellipsometry (A. Röseler, personal communication), from which the infrared cross-section in Fig. 3 is calculated. The extinction ΔC comes out much larger than with the low-absorbing materials Si and Au, and behaves like the known far-field absorption spectra. Within the range of available CO_2 laser lines, we chose the two marked in Fig. 3 for imaging. The composite sample was prepared by sequential drying of PS and PMMA drops on cleaved NaCl, then dissolving the NaCl in water to yield a nearly flat polymer surface with sharp material boundaries.

The infrared wavelength has a clear effect on the infrared image (Fig. 4, right-hand panels). As λ is tuned to a vibrational resonance of either material, the infrared image of the absorbing material is darker. This effect disappears at absorption-free wavelengths, giving clear evidence of near-field vibrational image contrast. The topography images (Fig. 4, left-hand panels) are reproduced well, and show an inclusion with its surface 7 nm below its surroundings. The inclusion can be assigned to be PS material because it shows a slightly different AFM tapping-phase contrast (not shown). The diagonal dark stripes are 3-nm-deep depressions in the surface, originating from crystal terraces in the cleaved NaCl surface. Their

absence in the infrared images again confirms the absence of topography-related artefacts.

The infrared material contrast in Fig. 4 amounts to 5% and 4%, respectively, at $\lambda = 9.68 \mu\text{m}$ and $10.17 \mu\text{m}$. This is a surprisingly large vibrational contrast even though it is significantly smaller than that calculated in Fig. 3, due to ΔC_b as stated above. To put this in perspective, we consider the bulk penetration depths for PMMA and PS at $\lambda = 10.17 \mu\text{m}$ which are 14 and $50 \mu\text{m}$, respectively. To observe a material contrast of 4% in a single-beam transmittance experiment, one would need a sample thickness $t \approx 0.04/(1/14 - 1/50) \mu\text{m} = 0.8 \mu\text{m}$, an interaction depth not available in the near-field-microscope due to its $\sim 50\text{-nm}$ spatial resolution. Here we have experimental evidence of enhanced infrared absorption mediated by the aperture-less near-field probe tip. Assuming the near-field interaction depth to be 50 nm, the observed enhancement factor is ~ 16 . The enhancement predicted from our model for a 30-nm-radius Au sphere (Fig. 3) is about 100 times higher than that, for a total of over three orders of magnitude. Enhanced electromagnetic interaction near a roughened or grainy metal surface is well known in Raman spectroscopy (surface-enhanced Raman scattering; SERS¹⁴) and less well known in infrared absorption (surface-enhanced infrared absorption; SEIRA⁷). Figure 4 is direct evidence of SEIRA initiated by a single metal particle, the gold tip of our microscope. It thus demonstrates how to probe SEIRA theory directly, without problems arising from size distribution and ensemble averaging.

Our experiment shows that infrared vibrational microscopy can be done at high resolution; even weak absorption provides sufficient contrast through an enhancement mechanism offered by the probing metal tip. With an infrared spectrum taken at each pixel one could envision chemical imaging with $<100\text{-nm}$ spatial resolution, corresponding to a material requirement of no more than $(100 \text{ nm})^3$, that is, 1 attolitre. □

Received 31 December 1998; accepted 9 March 1999.

1. Ash, E. A. & Nicholls, G. Super-resolution aperture scanning microscope. *Nature* **237**, 510–512 (1972).
2. Betzig, E., Trautmann, J. K., Harris, T. D., Weiner, J. S. & Kostelak, R. L. Breaking the diffraction barrier: optical microscopy on a nanometric scale. *Science* **251**, 1468–1470 (1991).
3. Wessel, J. Surface-enhanced optical microscopy. *J. Opt. Soc. Am. B* **2**, 1538–1540 (1985).
4. Zenhausern, F., Martin, Y. & Wickramasinghe, H. K. Scanning interferometric apertureless microscopy: optical imaging at 10 Ångström resolution. *Science* **269**, 1083–1085 (1995).
5. Lahrech, A., Bachelot, R., Gleyzes, P. & Boccara, A. C. Infrared-reflection-mode near-field microscopy using an apertureless probe with a resolution of $\lambda/600$. *Opt. Lett.* **21**, 1315–1317 (1996).
6. Knoll, B. & Keilmann, F. Scanning microscopy by mid-infrared near-field scattering. *Appl. Phys. A* **66**,

- 477–481 (1998).
7. Röseler, A. & Korte, E. H. Surface enhanced infrared absorption observed with attenuated total reflection (ATR-SEIRA): modeling the optical response. *Fresenius J. Anal. Chem.* **362**, 51–57 (1998).
 8. Lewis, A., Isaacson, M., Harootunian, A. & Muray, A. Development of a 500 Å spatial resolution light microscope. *Ultramicroscopy* **13**, 227–232 (1984).
 9. Pohl, D. W., Denk, W. & Lanz, M. Optical stethoscopy: Image recording with resolution $\lambda/20$. *Appl. Phys. Lett.* **44**, 651–653 (1984).
 10. Piednoir, A., Licoppe, C. & Creuzet, F. Imaging and local infrared-spectroscopy with a near-field optical microscope. *Opt. Commun.* **129**, 414–422 (1996).
 11. Knoll, B., Keilmann, F., Kramer, A. & Guckenberger, R. Contrast of microwave near-field microscopy. *Appl. Phys. Lett.* **70**, 2667–2669 (1997).
 12. Bohren, C. F. & Huffman, D. R. Absorption and scattering of light by small particles. (Wiley & Sons, New York, 1983).
 13. Völcker, M., Krieger, W. & Walther, H. Detection of local conductivity by laser-frequency mixing in a scanning force microscope. *J. Appl. Phys.* **74**, 5426–5431 (1993).
 14. Otto, A., Mrozek, I., Grabbhorn, H. & Akemann, W. Surface-enhanced Raman scattering. *J. Phys.: Condens. Matter* **4**, 1143–1212 (1992).

Acknowledgements. We thank H. Sturm for supplying samples and A. Röseler for supplying spectral data. Discussions with R. Guckenberger, A. Kramer, and D.v.d. Weide are acknowledged.

Correspondence and requests for materials should be addressed to F.K. (e-mail: keilmann@biochem.mpg.de)

Radon emanation and electric potential variations associated with transient deformation near reservoir lakes

M. Trique*, P. Richon†, F. Perrier*, J. P. Avouac* & J. C. Sabroux†

* Laboratoire de Détection et de Géophysique, DASE, Commissariat à l'Energie Atomique, BP 12, F-91680 Bruyères-le-Châtel, France

† Institut de Protection et de Sécurité Nucléaire, 60-68, BP 6, F-92265 Fontenay aux Roses, France

Two of the most often cited earthquake precursors are radon emanation and electric potential variations^{1–6}, but these few reported examples have generally been deemed questionable^{7–11}. If a mechanism relating crustal deformation to radon emanation or electrical signals does indeed exist, it is thought to involve fluids^{12–19}. Some preliminary insight has been gained into these processes from the study of natural systems under controlled mechanical and hydrological conditions²⁰. Here we report electric potential variations, radon emanation and deformation measurements recorded since 1995 in the French Alps in the vicinity of two artificial lakes which have strong seasonal variations in water level of more than 50 metres. We observe that electric potential variations and radon emanations are repeatedly associated with transient deformation events induced by variations in lake levels. These events are characterized by a change in ground tilt which deviates from the expected elastic response, and are associated with periods of accelerating strain, which suggests that accelerated loading can enhance fluid transport properties. Qualitatively, this behaviour can be accounted for by a model in which straining induces fluid overpressure and dynamic flow in cracks. These observations may shed light on the sensitivity of rock transport properties to deformation.

The site that we studied is located in the French Alps, at the geological contact between the Belledonne crystalline basement to the west, and highly tectonized Permo-Triassic sedimentary units to the east (Fig. 1). The levels of the lakes vary on a yearly cycle by ~70 m for the Roselend lake and 50 m for Gittaz lake (see Supplementary Information). A dead-end tunnel located 600 m northeast of the Roselend dam (Fig. 1) is equipped with two long-period seismometers (to measure the tilts in the north–south and east–west directions), one microbarograph and three radon detectors. An array of electrodes, composed of 14 measurement points, and a

meteorological station were set across the Sur-Frêres ridge which separates the two lakes (Fig. 1).

The experiment has been running since November 1995. The radon activity and the north–south tilt observed in the tunnel, the potential difference between the RH and EO points (see Fig. 1) and the Roselend lake level are shown as a function of time in Fig. 2. The radon activity in the inner room in the tunnel (see Methods) is characterized by a low background level of 650 Bq m⁻³, and by a number of bursts with amplitudes ranging from 2,000 to over 18,000 Bq m⁻³ and durations of one to nine weeks. The radon bursts with amplitudes greater than 4,000 Bq m⁻³ (labelled 2, 3, 4, 6, 8, 10, 11 and 13 in Fig. 2) are observed to correlate systematically with decreases of the RH-EO potential difference.

The electric potential variations are occurring between four days before and one day after the radon bursts. Their amplitudes range from –6 to –20 mV for time durations varying between 5 and 33 d. They do not scale in magnitude and duration time with the radon bursts. As these potential variations are not observed on the other dipoles connected to the point EO, they must originate at the point RH. Given that the observed transient electric potential variations

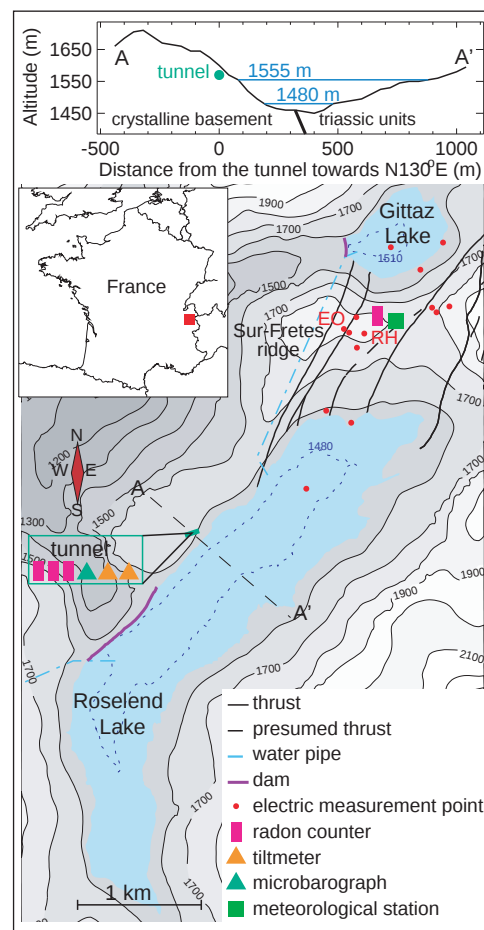


Figure 1 Layout of the Sur-Frêres experiment. The general location of the experiment is indicated by the red square in the inset to the lower figure; in the lower figure, the blue dashed curve in each lake shows the extents of the lakes at their lowest level. The topographic profile along the segment A–A' is represented in the top figure. The Roselend lake has a capacity of 187×10^6 m³ of water. The smaller Gittaz lake has a capacity of 13×10^6 m³ and communicates with the Roselend lake through an underground water pipe (blue dashed curve). The lakes are enclosed by two dams set against the crystalline rocks (purple). The lake level cycle is strongly correlated to the yearly meteorological cycle only during the snow melting phase in spring and summer. The lakes are emptied according to the needs of electric power production by a pipe (blue dashed curve) linking the Roselend lake to a power station located 15 km west.

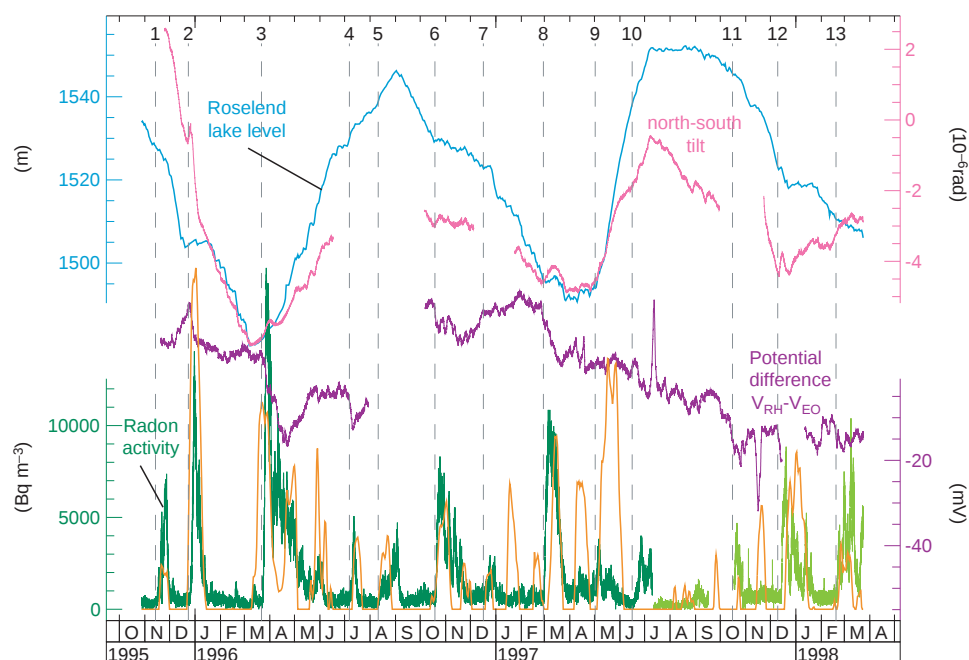


Figure 2 Temporal variations of Roselend lake level (blue), north-south tilt (red), electric potential (purple), and radon activity (green). The interruption in the series are due to storms and work on the power line. The radon activity was measured in the inner room (see Methods) from November 1995 to July 1997 (dark green), and in the corridor from July 1997 to April 1998 (light green). The north-south tilt and

electric potential data segments are arbitrarily shifted from each other. The electrical data are low-pass-filtered to remove periods <2 d, and their average value is arbitrary. The orange curve is the positive part of the discrete second derivative applied on the lake level data and expressed in arbitrary units. The lake-level data are provided by Electricité De France.

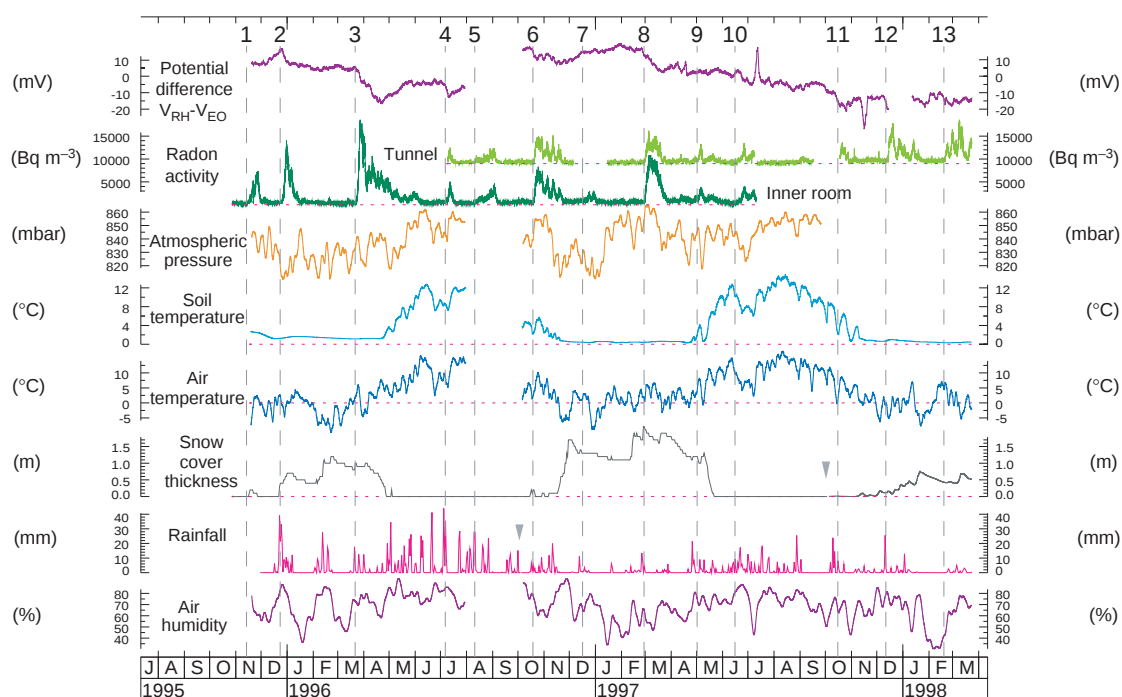


Figure 3 Potential difference of dipole RH-EO, radon activity measurements and meteorological parameters as a function of time. The meteorological parameters are recorded with a sampling interval of 5 min. The electric potential, soil temperature, atmospheric pressure, air humidity, and air temperature data are low-pass-filtered to remove the periods <2 d. The dark green curve corresponds to the autonomous radon probe located in the inner room of the tunnel, whereas the light green curve corresponds to one of the two radon probes located in the

tunnel. The grey reversed triangles indicate when the data were first acquired by the local meteorological devices. From November 1995 to October 1997, the snow height data, provided by Electricité De France, were recorded by a snowmeter located 2 km away and 150 m higher from the Sur-Frères ridge. From October 1995 to October 1996, the rainfall data given by Météo France were acquired at Luce station located 6 km away and 800 m below the site.

do not present any pattern similar to the soil temperature data, we can exclude an electrode response to soil temperature gradients or soil humidity changes²¹. Transient electric potential variations can also be generated by the shallow water circulation²¹ systematically induced by rainfall or snow melting. A direct systematic and trivial physical relation between the rain or snow melting events and the observed electric potential variations (see Methods and Supplementary Information), and more generally with any meteorological parameters, can be excluded (Fig. 3). Also, the radon bursts are not instrumental artefacts: they are recorded by independent sensors in the tunnel (Fig. 3). The radon bursts observed in the tunnel have no systematic correlation with any local meteorological parameters (Fig. 3) which could influence the radon activity in the air of the tunnel. In particular, neither the “pumping” effect²² due to decreasing barometric pressure, nor thermal circulation through the temperature gradient between the outside and the inside of the tunnel, can explain the bursts. Other effects, such as air humidity and high-speed winds, are known to affect radon emanation²², but these can be excluded in our case; we can also exclude the blocking of the tunnel entrance by snow (Fig. 3). Rainfall, snowfall and snow melting could also affect the radon activity in the tunnel because such processes could modify the water saturation around it: but because no systematic relation appears between the signals (see Methods and Supplementary Information), this hypothesis is excluded (Fig. 3). The radon probe located on the Sur-Frères ridge did not record any similar radon bursts. The radon bursts measured in the tunnel are therefore not caused by meteorological effects.

Because the radon level in the tunnel is controlled by the radon emanation and the transport paths of the rock matrix, the variations of the radon level measured in the tunnel could be due to changes in the radon transport properties. The emanation of the rock matrix has been measured in the laboratory for dry gneiss samples, and amounts to $8.3 \text{ atoms m}^{-2} \text{ s}^{-1}$. The background radon level is therefore consistent with radon drained at the tunnel walls from a 4-cm-thick zone of saturated rocks. The radon bursts would imply either a transient change of rock saturation, from saturated to dry, or an increase of the rock volume from which the radon is drained. In the latter case, an increase of the volume of drained zone by a factor of ~ 100 is necessary to account for the radon peak levels.

The radon bursts correlate with positive-slope kinks in the plot of lake level against time (Fig. 2). We treat these kinks as follows: from the lake-level data series $h(t)$, a discrete second derivative function is built as $f(i) = h(i) - 2h(i - T) + h(i - 2T)$, where T is some fixed time delay. For a time delay of $T = 14 \text{ d}$, which was found to be most appropriate, and keeping only positive values of $f(t)$, the variations of $f(t)$ reproduce well the observed radon bursts (Fig. 2). Indeed, given a time delay between the signals ranging between -10 and 10 days, 63% of the positive variations of $f(t)$ (out of 16 events) correlate to the radon bursts (see Supplementary Information). The score is only 27% on average if the radon bursts are redistributed randomly. The proportion of the positive variations of $f(t)$ which could be related to the transient electrical variations, assuming the same time delay, amounts to 53% out of 15 events. A lesser value (30% with the same time delay) is obtained for a random time series of electrical variations. Thus, the electric and radon signals both tend to be produced by a mechanical effect, namely the accelerated straining of the medium.

This is further supported by the tilt-meter data. The tilt-meter recorded a yearly variation which amounts to several microradians and which is linearly correlated to the lake level (Fig. 2). The measured linear factor of $\sim 0.10 \mu\text{rad m}^{-1}$ is compatible with the elastic deformation due to the lake loading. Indeed, using the values of the Lamé parameters deduced from seismic sounding²³, the expected north-south tilt per metre of water amounts to $0.090 \mu\text{rad m}^{-1}$, in agreement with the observed value. The tilt-meter response is thus observed to depart from the average elastic response during the radon and electric potential anomalies. This can be clearly seen in Fig. 4, where expanded-scale views around two radon bursts (events 3 and 8) are presented. The tilt response is proportional to the lake level for constant loading velocity, but during the radon burst and coeval with the change in the potential difference, a response amplified by a factor of 3 to 5 is observed on the north-south tilt compared with the lake level.

In principle, transient tilt deformations could be generated by meteorological events; for example, by the cracks filling with water supplied by the rainfall or snow melting. However, such an artefact would appear systematically, and would also be observed with a larger amplitude at the snow-melting time. In addition, there would be no relationship between the time structure of the anomaly (basically dominated by the crack permeability) and the time structure of the lake level. We therefore consider that the tilt anomalies are not a meteorological artefact but are rather a consequence of a different mechanical response.

The time delay of 14 d required for the discrete second derivative to match the observed radon bursts (Fig. 2) suggests that the pore water may play a role. Indeed, according to the permeability (10^{-16} m^2) measured *in situ*²³, the diffusion time for a lake-level change to affect the tunnel is estimated²⁴ to range from 1 to 30 d. The different mechanical response observed for perturbations with a timescale smaller than 14 d suggests that the response then occurs in the undrained mode. However, the difference in the mechanical response between the drained and undrained conditions can at most amount to 30% (ref. 24). The observed amplification (of a factor 4 on average) must then result from another mechanism.

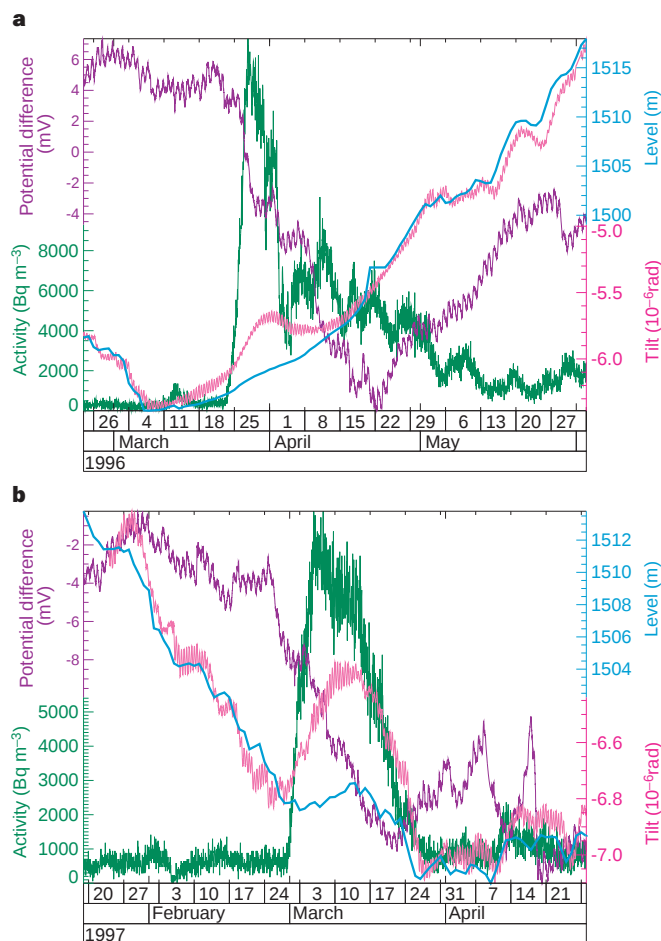


Figure 4 Roselend lake level (blue), north-south tilt (red), radon activity (green) and electric potential $V_{RH} - V_{EO}$ (purple). Data are shown for the time interval encompassing events 3 (a) and 8 (b).

The observed response may be expected from the following model. We consider a medium consisting of a low-permeability matrix with fluid-filled cracks. This model is analogous to that used to calculate the groundwater outflow in response to coseismic deformation¹⁵. If a strain rate de/dt is applied to the medium (compressive strain is considered as positive), the variation of pore volume of the cracks induces a dynamic equilibrium, in which the pore-pressure variation is balanced by the fluid flow out of the rock matrix. A first-order approximation of the pore pressure p leads to the following relationship: $p = (\eta L^2 r/k)(de/dt)$, where η is the fluid viscosity, L the characteristic length along percolation paths, k the permeability of the medium, and r the aspect ratio of the cracks. This relationship implies a pore-pressure increase (in response to accelerated lake loading) that could reach 0.1 to 1 bar, assuming a typical percolation length scale of 1 to 10 m, a permeability of crystalline rocks²³ of 10^{-16} m^2 , an aspect ratio of 10^3 for the cracks, and a strain rate of 10^{-7} d^{-1} . Such overpressures may be sufficient to produce local microfracturing, and hence an increase of transport properties or the exposure of new fracture surfaces. The rapid transport to the tunnel atmosphere of the previously trapped radon gas could then account for the observed radon bursts.

The observed electric potential variations could also be produced through the electrokinetic effect by a similar enhancement of the rock transmissivity. The magnitude of the pore pressure of the fluids involved in the electrical effects can be estimated from the data of this experiment²⁵. Indeed, the variations of the potential of one point of the electrical array are linearly correlated to the Roselend lake level, with a correlation of 20 mV per 0.1 MPa (ref. 25). Using this value, the potential variations (of the order of 10 mV) observed on the RH-EO dipole could result from the connection of two systems with a 0.5-bar pore-pressure difference. This order of magnitude could be expected in the presence of an aquifer with a piezometric level higher than the mean level in the surrounding medium. Indeed, such a perched aquifer could be related to a 60-m-thick low-resistivity zone below point RH observed in a very-low-frequency electromagnetic survey performed on the Sur-Frères ridge in July 1995 (S. Hautot and P. Tarits, personal communication).

The present experiment is not the first observation of an anelastic response to small changes in loading rate in a natural system. For example, induced seismicity near the Nurek reservoir, Tadjikistan²⁶, and the Koyna reservoir, India²⁷, appears to be controlled by the rate of change in the water level. Moreover, transient permeability enhancements have also been proposed to explain the occurrence of hydrothermal precursors to large earthquakes in northern California²⁸ and changes in hydrology occurring after large earthquakes²⁹. The physical processes involved in the signals recorded at the Sur-Frères site may provide the basic understanding of the transport properties of kilometre-sized systems subject to stress variations: they may also be of interest for the monitoring of ground water or of waste-disposal sites. Although the applicability of our observations at the scale of active tectonic areas needs to be carefully considered, they provide an analogue to earthquake precursors, with the advantage that the mechanical forcing is known, as are the main characteristics of the medium. □

Methods

Tunnel characteristics. The nearly horizontal tunnel, drilled in gneiss, is 142 m long with a 2 m diameter, and contains an inner room of $4 \times 5 \times 2 \text{ m}$, located 60 m from the entrance and 40 m below the surface. This inner room is separated from the corridor by a metal door and the tunnel is closed by a second metal door. The inner room is not significantly ventilated, as the pressure difference between the inner room and outside has been measured to be $< 1 \text{ Pa}$. The radon background level recorded in the inner room does not show the typical seasonal variation due to air exchange driven by temperature difference between the inside and the outside, as is generally observed in caves²², attesting

again to a poor ventilation of the inner room. The tunnel closing was reinforced in October 1997 by two air-tight plastic doors to further reduce ventilation. The radon background level recorded in the tunnel rose from 250 to 650 Bq m^{-3} , as measured in the inner room, suggesting an enhancement of the air confinement in the tunnel. The air relative humidity in the tunnel is close to 100%. The temperature is stable at $6.6 \pm 0.2^\circ \text{C}$ over the whole year.

Tilt measurements. The tiltmeters have a resolution of $20 \times 10^{-9} \text{ rad}$. They have a pressure sensitivity (determined in the laboratory) of 0.012 and $0.006 \mu\text{rad mbar}^{-1}$ for the north-south and east-west tilt-meters, respectively. The measurements are corrected for pressure variations using a microbarograph, having an accuracy of $5 \times 10^{-2} \text{ mbar}$. The tilt data are filtered to remove frequencies higher than 1 Hz, and averaged at one point per minute. The east-west tiltmeter was affected by instrumental long-term drift and therefore is not used in the study.

Radon measurements. The radon probes are commercially available (as BARASOL). They consist of a cylindrical measurement chamber in equilibrium with the outside air through a diffusion filter. The decay of ^{222}Rn is detected by a solid-state silicon detector. The counts are integrated over an hour. The sensitivity of the device is 50 Bq m^{-3} (1 count per hour). One autonomous probe was installed in the inner room in November 1995, and two other probes connected to a modem were added in the corridor in August 1996. One autonomous radon probe was also set in the soil at the Sur-Frères ridge (Fig. 1) in November 1995.

Electrical measurements. We used second generation Pb/PbCl₂/kaolinite Petiau³⁰ unpolarizable electrodes. These electrodes are installed in 1.2 to 1.5 m deep, 40 cm diameter holes in 30 l of salted clay. This set-up was tested during an international comparison experiment³⁰, and was demonstrated to provide a stability of the order of 1 mV per year with a drift noise between 0.4 and 0.8 mV per month. The electric potential array is composed of 14 measurement points. The electrodes are combined into 20 dipoles in order to provide as much redundancy as possible. They are connected to a measuring station located on the ridge, through buried or aerial cables. The potential differences are recorded with a sampling rate of one per minute.

Correlation tests. The proportion of the rainfall, snowfall and snow-melting events which could be related to the radon bursts, assuming a maximal time delay of 10 d, amounts to 12% (out of 65 events; see Supplementary Information). For comparison, a similar value of 14% is obtained on average if the radon bursts are redistributed randomly. Therefore, no systematic relationship appears between these meteorological events and the radon bursts. Similarly, the transient electric potential variations are not systematically related to rainfall or snow melting. The proportion of these meteorological events which could be related to the transient electrical variations, assuming a maximal time delay of 10 d, amounts to 20% out of 42 events. For comparison, a similar value of 15% is obtained on average, if the electric signals are redistributed randomly.

Radon emanation in the laboratory. The equilibrium radon activity measurement performed on dry gneiss samples of $6 \times 10^{-5} \text{ m}^3$, using a measuring chamber having a free volume of $9 \times 10^{-4} \text{ m}^3$, amounts to 155 Bq m^{-3} . The radon emanation value of $8.3 \text{ atoms m}^{-2} \text{ s}^{-1}$ mentioned in the text was deduced without taking into account the internal porous surface of the sample.

Received 13 August 1998; accepted 4 March 1999.

1. Wakita, H., Nakamura, Y. & Sano, Y. Short-term and intermediate-term geochemical precursors. *Pure Appl. Geophys.* **126**, 267–278 (1988).
2. Igarashi, G. *et al.* Ground-water radon anomaly before the Kobe earthquake in Japan. *Science* **269**, 60–61 (1995).
3. Virk, H. S. & Singh, B. Radon recording of Uttarkashi earthquake. *Geophys. Res. Lett.* **21**, 737–740 (1994).
4. Raleigh, B. *et al.* The prediction of the Haicheng earthquake. *Eos* **58**, 236–272 (1977).
5. Corwin, R. F. & Morrison, H. F. Self-potential variations preceding earthquakes in Central California. *Geophys. Res. Lett.* **4**, 171–174 (1977).
6. Varotsos, P., Alexopoulos, K. & Lazaridou, M. Latest aspects of earthquake prediction in Greece by seismic electric signals, II. *Tectonophysics* **224**, 1–37 (1993).
7. Geller, R. J. (ed.) *Geophys. Res. Lett.* **23** (debate on “VAN” special issue) 1291–1452 (1996).
8. Pham, V. N., Boyer, D., Chouliaras, G., Le Mouél, J. L. & Rossignol, J. C. Characteristics of electromagnetic noise in the Ioannina region (Greece): a possible origin for so called “Seismic Electric Signal” (SES). *Geophys. Res. Lett.* **25**, 2229–2232 (1998).
9. Morrison, H. F., Fernandez, R. & Corwin, R. F. Earth resistivity, self potential variations, and earthquakes: a negative result for $M = 4.0$. *Geophys. Res. Lett.* **6**, 139–142 (1979).
10. Geller, R. J. Earthquake prediction: a critical review. *Geophys. J. Int.* **131**, 425–450 (1997).
11. Wyss, M. Second round of evaluations of proposed earthquake precursors. *Pure Appl. Geophys.* **149**, 3–16 (1997).

12. Scholz, C. H., Sykes, L. R. & Aggarwal, Y. P. Earthquake prediction: a physical basis. *Science* **181**, 803–810 (1973).
13. Igarashi, G., Wakita, H. & Sato, T. Precursory and coseismic anomalies in well water levels observed for the February 2, 1992 Tokyo Bay earthquake. *Geophys. Res. Lett.* **19**, 1583–1586 (1992).
14. Roeloffs, E. A. Hydrologic precursors to earthquakes: a review. *Pure Appl. Geophys.* **126**, 177–209 (1988).
15. Muir-Wood, R. & King, G. C. P. Hydrological signatures of earthquake strain. *J. Geophys. Res.* **98**, 22035–22068 (1993).
16. Mizutani, H., Ishido, T., Yokotura, T. & Onishi, S. Electrokinetic phenomena associated with earthquakes. *Geophys. Res. Lett.* **3**, 365–368 (1976).
17. Jouriaux, L. & Pozzi, J. P. Streaming potential and permeability of saturated sandstones under triaxial stress: consequences for electrotelluric anomalies prior to earthquakes. *J. Geophys. Res.* **100**, 10197–10209 (1995).
18. Bernard, P. Plausibility of long distance electrotelluric precursors to earthquakes. *J. Geophys. Res.* **97**, 17531–17546 (1992).
19. King, C. Y. Gas geochemistry applied to earthquake prediction: an overview. *J. Geophys. Res.* **91**, 12269–12281 (1986).
20. Morat, P. & Le Mouél, J. L. Electrical signals generated by the collapse of the pillars of a gypsum quarry. *C.R. Acad. Sci.* **308**, 33–38 (1989).
21. Perrier, F. et al. A one-year systematic study of electrodes for long period measurement of the electric field in geophysical environments. *J. Geomagn. Geoelectr.* **49**, 1677–1696 (1997).
22. Wilkening, M. H. & Watkins, D. E. Air exchange and ^{222}Rn concentrations in the Carlsbad caverns. *Health Phys.* **31**, 139–145 (1976).
23. Direction de l'Équipement Électricité De France, Note de synthèse géologique et technique, *Barrage de Roselend, Fondation de l'Ouvrage 1–12*, (Électricité De France, Aix-en-Provence, 1952–1962). (In French.)
24. Roeloffs, E. A. Fault stability changes induced beneath a reservoir with cyclic variations in water level. *J. Geophys. Res.* **93**, 2107–2124 (1988).
25. Perrier, F., Trique, M., Hautot, S., Avouac, J. P. & Tarits, P. Electrical variations associated with yearly lake level variations. *Geophys. Res. Lett.* **25**, 1955–1958 (1998).
26. Simpson, D. W. & Negmatullaev, S. K. Induced seismicity at Nurek Reservoir, Tadzhikistan, USSR. *Bull. Seismol. Soc. Am.* **71**, 1561–1586 (1981).
27. Gupta, H. K. Induced seismicity hazard mitigation through water level manipulation at Koyna, India: a suggestion. *Bull. Seismol. Soc. Am.* **73**, 679–682 (1983).
28. Silver, P. G. & Valette-Silver, N. J. Detection of hydrothermal precursors to large Northern California earthquakes. *Science* **257**, 1363–1367 (1992).
29. Rojstaczer, S., Wolf, S. & Michel, R. Permeability enhancement in the shallow crust as a cause of earthquake-induced hydrological processes. *Nature* **373**, 237–239 (1995).
30. Petiau, G. In *Proc. of the Workshop "Electrodes"* (eds Clerc, G., Perrier, F., Petiau, G. & Menville, M.) (Report, Centre de Recherches Géophysiques de Garchy, 1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank Y. Caristan, A. Meesters and R. Pican for support, and EDF for its water level data and access to the lakes. We also thank the members of the technical team from the Département Analyse et Surveillance de l'Environnement, CEA, for their dedication in performing their work in often difficult weather conditions.

Correspondence and requests for materials should be addressed to M.T. (e-mail: trique@ldg.bruyeres.cea.fr).

Three-dimensional preservation of foot movements in Triassic theropod dinosaurs

Stephen M. Gatesy*, Kevin M. Middleton*, Farish A. Jenkins Jr† & Neil H. Shubin‡

* Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island 02912, USA

† Department of Organismic and Evolutionary Biology, and Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA

‡ Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Dinosaur footprints have been used extensively as biostratigraphic markers, environmental indicators, measures of faunal diversity and evidence of group behaviour^{1–5}. Trackways have also been used to estimate locomotor posture, gait and speed^{6–11}, but most prints, being shallow impressions of a foot's plantar surface, provide little evidence of the details of limb excursion. Here we describe Late Triassic trackways from East Greenland, made by theropods walking on substrates of different consistency and sinking to variable depths, that preserve three-dimensional records of foot movement. Triassic theropod prints share many features with those of ground-dwelling birds, but also demonstrate significant functional differences in position of the hallux (digit I), foot posture and hindlimb excursion.

Dinosaur trackways are common in the Ørsted Dal Member of

the uppermost Fleming Fjord Formation (Norian-Rhaetian) of Jameson Land, East Greenland¹². These strata consist of cyclically bedded fine-grained siliciclastic and carbonate-bearing units laid down as part of a large rift lake system¹³. Although the footprints are structurally diverse, almost all appear to be members of a gradational series (Fig. 1a–f). At one extreme are clearly preserved tridactyl tracks referable to the ichnogenus *Grallator* (*Anchisauripus*)^{1,14,15}. Such prints are commonly attributed to bipedal theropods¹⁶ and preserve impressions of digital pads, claws and, in some cases, skin (Fig. 1a). At the other extreme are tetradactyl, elongate prints in which the digital impressions are smooth channels constricting to narrow slits (Fig. 1f). The apparent impression of digit III is markedly longer than those of other digits; its distal end is elevated and expanded to resemble a fumarole-like crater. A posteromedially directed imprint of digit I is always present adjacent to an impression of the metatarsus. Most of the Ørsted Dal prints are intermediate in form between these two extremes (Fig. 1b–e).

The differences between tridactyl (Fig. 1a) and tetradactyl (Fig. 1f) tracks are comparable to those that have been used to distinguish faunal constituents and define ichnotaxa^{1,2}. However, two lines of evidence support the hypothesis that track variation is related to theropods traversing substrates of variable consistency, rather than to differences in foot morphology. First, some individual trackways exhibit more than one variant, although no single trackway preserves a complete gradational series. Second, guinea-fowl and turkeys walking on substrates of different consistencies closely duplicate the structural continuum of the Greenlandic prints (Fig. 1g–j). Whereas shallow prints on a relatively firm substrate accurately replicate the plantar form of these birds (Fig. 1g), successively deeper prints into increasingly wetter sediments record the entry of the initially splayed foot and the subsequent digital convergence as the foot is extracted (Figs 1h–j, 2a). Footprints in muds of intermediate consistency show separate entry and exit furrows for digits II and IV (Fig. 1h, i). In the deepest, wettest muds, the entire foot (with digits I–IV convergent) is extracted at the anterior end of the metatarsal-digit III furrow, raising a mound of sediment (Figs 1j, 2b). Thus, as mud hydration and penetration depth increase, bird tracks elongate because the substrate intercepts foot movements that normally occur above the surface. In the elongate Triassic footprints, the subsurface pathway of the toes disrupts sedimentary laminae, which can be seen in serial sections. These three-dimensional records confirm that, as in living birds, early Mesozoic theropods sank down and forward, and extracted the foot with convergent toes. Although evidence for slight digital approximation at the end of the stance phase has been reported from shallower tracks¹⁷, the elongate Greenlandic tracks document complete toe convergence, which is retained as a primitive feature among birds.

Although the general correspondence between extant avian and Greenlandic trackways reflects some pedal similarities, there are important functional differences between basal theropods and birds. The hallux (digit I) in ceratosaurs, such as *Coelophysis*, is sufficiently shorter than the other digits that on firm substrates it does not contact the ground (cf. Fig. 1a, b). The hallucal metatarsal, applied as a flattened splint to the midshaft of metatarsal II (Fig. 3), bears an asymmetrical condyle that permits a range of about 70° of flexion/extension and about 25° of ab/adduction of the metatarsophalangeal joint. The single interphalangeal joint, which is uniaxial, permits only flexion/extension. Fixation of metatarsal I and the configuration of the metatarsophalangeal and interphalangeal joints preclude retroversion. No known Late Triassic–Early Jurassic theropod shows evidence of an avian-like reversed hallux (*Eoraptor*, *Herrerasaurus*, *Coelophysis*, *Syntarsus*, *Procompsognathus*, *Segisaurus*, *Dilophosaurus*^{18–24}). Nonetheless, the posteromedially oriented hallucal print of some tracks, which appears superficially comparable to the impression of the posteriorly directed hallux of birds (compare

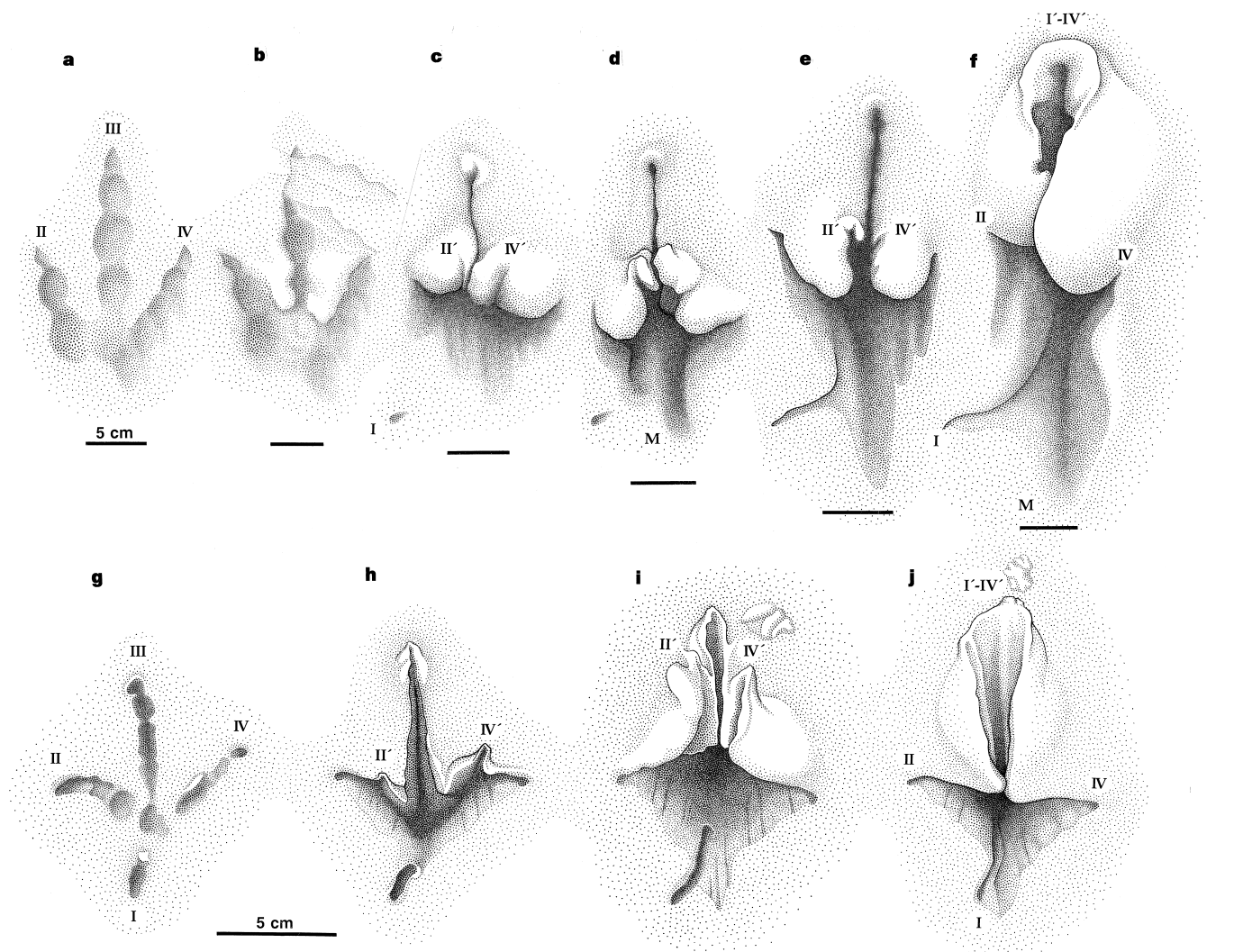


Figure 1 The spectrum of theropod tracks preserved in the Triassic Fleming Fjord Formation of Greenland (**a–f**) is comparable to that produced by an extant bird (the helmeted guineafowl, *Numida meleagris*) (**g–j**). On firm substrates, tracks closely reflect plantar morphology (**a, g**). The feet sink progressively deeper as sediment hydration increases (**b–f** and **h–j**). As a result, digits II and IV begin to converge before the foot is withdrawn, thereby creating separate entry (II, IV) and exit (II', IV') furrows (**c–e, h, i**) or a single terminal exit (**f, j**). Unlike birds, however,

Triassic theropods left a distinct metatarsal impression (M) in deeper tracks (**d–f**). Greenlandic footprints were drawn from separate trackways and represent different individuals as well as variant details, whereas the avian prints were made by the same bird. The track illustrated in **b** is partially obscured by an overprint of another track. Theropod tracks (**a–f**) are registered in the Geological Museum, University of Copenhagen, as MGUH VP 3386–3391.

Fig. 1c–f and g–j), has led to speculation that undiscovered theropods with a reversed hallux existed in the early Mesozoic^{2,25–27}.

We propose that the Greenlandic trackways could have been created by ceratosaurs in which the hallux was slightly abducted and flexed, but not reversed. In prints of intermediate depth, contact of the hallux with the substrate was transient, leaving an isolated mark posteromedial to the main track (Fig. 1c, d). On more pliable muds, the hallux followed the forward and downward trajectory of the foot (~45°) into the sediment. With the hallux oriented more vertically than the angle of foot penetration, an oblique impression was produced as the digit incised the surface in an anterolateral direction (Figs 1e, f, 3). Under these conditions, a foot such as that of *Coelophysis*, with an anteromedially oriented digit I, could produce a posteromedially directed hallucal furrow. Equating the hallucal entry furrow with the anatomical orientation of this digit accounts for the erroneous interpretation of a reversed hallux among early Mesozoic theropod trackmakers.

A second distinction between the Greenlandic tracks and those of living birds is the metatarsal imprint. The deepest theropod tracks are posteriorly elongated by a smooth metatarsal impression with a

shallow slope of 25–30° (Fig. 1e, f). Turkeys and guineafowl, in contrast, leave little or no metatarsal print, even in the deepest tracks (Fig. 1i, j). As the avian foot penetrates down and forward, the metatarsus rapidly becomes more vertical than the angle of foot penetration, and thus no posterior metatarsal impression is created (Fig. 2b). In contrast, the distinct metatarsal imprint in Greenlandic tracks indicates that the initial foot posture was maintained during the early part of the stance phase in these theropods (Fig. 3).

The contrast between Triassic theropods and extant birds in the timing of metatarsal rotation implies differences in the movement of more proximal hindlimb segments. Among birds today, the stance phase of walking is characterized by large excursions at the knee and metatarsophalangeal joints, but little movement at the hip and ankle. During the first half of the stance phase, a relatively small femoral excursion is coupled with early elevation of the metatarsus. In contrast, walking crocodilians employ little movement at the knee and metatarsophalangeal joints; the metatarsus remains stable in early stance, when hip extension and ankle dorsiflexion predominate²⁸. On the basis of caudopelvic anatomy, retraction of the femur (hip extension) has been proposed as the principal

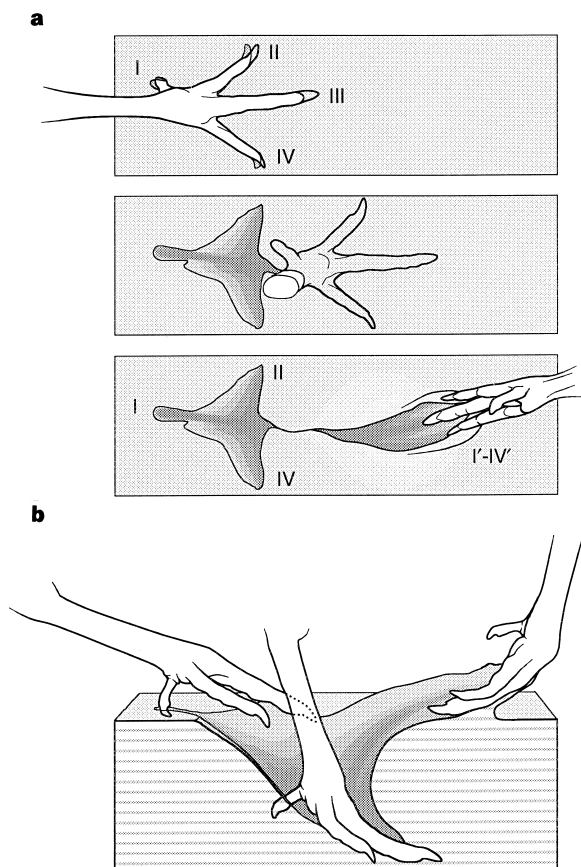


Figure 2 Early, middle, and late phases in the formation of a turkey footprint in deep mud. **a**, Dorsal and **b**, Lateral views. As the foot penetrates down and forward, the metatarsus becomes more vertical, leaving little or no impression at the rear of the track. The digits converge subsurface, emerging from the substrate (I'–IV') well in front of their points of entry (I–IV). Foot positions were reconstructed from video records of a single stride taken simultaneously in lateral and posterior views, and aligned with respect to the resulting track.

stance-phase movement in basal theropods^{29,30}. The impressions left by stable metatarsi in deep tracks from the Greenlandic Triassic are independent evidence that these theropods powered the early stance phase by femoral retraction, rather than by knee flexion as in living birds. □

Methods

Triassic theropod footprints. Fossil footprints, collected at 71°24.88' N, 22°33.17' W on the eastern slope of Wood Bjerg, Jameson Land, East Greenland, were mechanically prepared to remove infilled matrix. Four elongate prints were transversely or sagittally sectioned with a rotary rock saw at 5 mm intervals to trace the subsurface path of each toe and determine foot penetration angles. We assessed the mobility of the hallux in two specimens of *Coelophysis bauri* (Museum of Comparative Zoology 4334, American Museum of Natural History 7226) by mounting the digital elements forming the metatarsophalangeal and interphalangeal joints at the extremes of their excursion ranges as indicated by the geometry of the articular facets. Images of the maximal excursion ranges, taken in standard anatomical perspectives, were captured by a Toshiba Image Master CCD Color Camera mounted on a Nikon SMZ-2T binocular microscope, and printed on a Sony Mavigraph colour video printer. Three-dimensional reconstructions of foot movement were modelled (based on the pedal anatomy of *Coelophysis bauri*) and rendered using Alias/Wavefront Studio 8.5 on a Silicon Graphics Indigo² IMPACT computer. Foot positions were reconstructed on the basis of penetration angles, hallux cleft orientations, metatarsal impression slopes and toe convergence preserved in deep Greenlandic tracks.

Avian footprints. We videotaped an adult female turkey (*Meleagris gallopavo*)

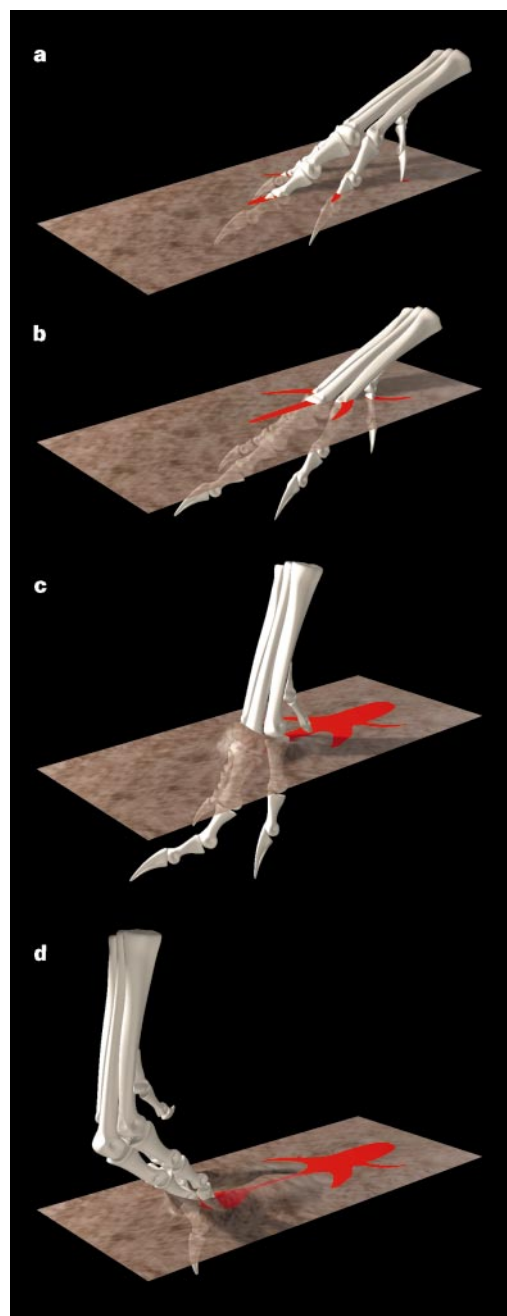


Figure 3 Three-dimensional computer reconstruction of theropod foot movements producing an elongate print in deep mud. Penetration through the substrate surface is shown in red. The hallux, which is not 'reversed' as in modern birds, leaves a posteromedially oriented entry furrow (**a**, **b**) as the digit plunges down and forward. The ankle is not immediately lifted during the stance phase, and as a result a metatarsal impression is created (**c**), elongating the track posteriorly. All digits converge beneath the surface and emerge together from the anterior end of the track (**d**).

and an adult male helmeted guineafowl (*Numida meleagris*) in simultaneous lateral and anterior/posterior views as they walked and ran across a clay-rich, sandy soil. The soil was variously hydrated to produce substrate consistencies ranging from a firm, plasticine-like surface, with a depth of 7 mm (Fig. 1g), to a highly compliant mud with a depth of 61 mm (Fig. 1j). Trackways and individual tracks were documented by stereophotography; some prints were cast in plaster.

Received 9 February; accepted 15 March 1999.

1. Hitchcock, E. *Ichnology of New England: A Report on the Sandstone of the Connecticut Valley, especially its Fossil Footmarks, made to the Government of the Commonwealth of Massachusetts* (W. White, Boston, 1858).

2. Lull, R. S. Triassic life of the Connecticut Valley. *Conn. State Geol. Nat. Hist. Surv. Bull.* **81**, 1–336 (1953).
3. Gillette, D. D. & Lockley, M. G. eds *Dinosaur Tracks and Traces* (Cambridge University Press, Cambridge, 1989).
4. Thulborn, T. *Dinosaur Tracks* (Chapman and Hall, London, 1990).
5. Lockley, M. G. *Tracking Dinosaurs: A New Look at an Ancient World* (Cambridge University Press, Cambridge, 1991).
6. Alexander, R. McN. Estimates of speeds of dinosaurs. *Nature* **261**, 129–130 (1976).
7. Coombs, W. P. Jr Swimming ability of carnivorous dinosaurs. *Science* **207**, 1198–1200 (1980).
8. Farlow, J. O. Estimates of dinosaur speeds from a new trackway site in Texas. *Nature* **294**, 747–748 (1981).
9. Thulborn, R. A. Speeds and gaits of dinosaurs. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **38**, 227–256 (1982).
10. Thulborn, R. A. & Wade, M. Dinosaur trackways in the Winton Formation (mid-Cretaceous) of Queensland. *Mem. Qd. Mus.* **21**, 413–517 (1984).
11. Padian, K. & Olsen, P. E. in *Dinosaur Tracks and Traces* (eds Gillette, D. D. & Lockley, M. G.) 231–241 (Cambridge University Press, Cambridge, 1989).
12. Jenkins, F. A. Jr et al. Late Triassic continental vertebrates and depositional environments of the Fleming Fjord Formation, Jameson Land, East Greenland. *Meddr. Grønland Geoscience* **32**, 1–25 (1994).
13. Clemmensen, L. B., Kent, D. V. & Jenkins, F. A. Jr A Late Triassic lake system in East Greenland: facies, depositional cycles and palaeoclimate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **140**, 135–159 (1998).
14. Lull, R. S. Fossil footprints of the Jura-Trias of North America. *Mem. Bos. Soc. Nat. Hist.* **1904**, 461–557 (1904).
15. Olsen, P. E. & Galton, P. M. A review of the reptile and amphibian assemblages from the Stormberg of southern Africa, with special emphasis on the footprints and the age of the Stormberg. *Palaeont. Afr.* **25**, 87–110 (1984).
16. Olsen, P. E., Smith, J. B. & McDonald, N. G. Type material of the type species of the classic theropod footprint genera *Eubrontes*, *Anchisauripus*, and *Grallator* (Early Jurassic, Hartford and Deerfield Basins, Connecticut and Massachusetts, U.S.A.). *J. Vert. Paleol.* **18**, 586–601 (1998).
17. Baird, D. Triassic reptile footprint faunules from Milford, New Jersey. *Bull. Mus. Comp. Zool.* **117**, 449–520 (1957).
18. Sereno, P. C., Forster, C. A., Rogers, R. R. & Monetta, A. M. Primitive dinosaur skeleton from Argentina and the early evolution of Dinosauria. *Nature* **361**, 64–66 (1993).
19. Novas, F. E. New information on the systematics and postcranial skeleton of *Herrerasaurus ischigualastensis* (Theropoda: Herrerasauridae) from the Ischigualasto Formation (Upper Triassic) of Argentina. *J. Vert. Paleol.* **13**, 400–423 (1993).
20. Colbert, E. H. The Triassic dinosaur *Coelophysis*. *Mus. N. Ariz. Bull.* **57**, 1–160 (1989).
21. Raath, M. A. A new coelurosaurian dinosaur from the Forest Sandstone of Rhodesia. *Arnoldia* **4**, 1–25 (1969).
22. Sereno, P. C. & Wild, R. *Procompsognathus*: theropod, “thecodont” or both? *J. Vert. Paleol.* **12**, 435–458 (1992).
23. Camp, C. L. A new type of small bipedal dinosaur from the Navajo Sandstone of Arizona. *Univ. Cal. Pub. Geol. Sci.* **24**, 39–56 (1936).
24. Welles, S. P. *Dilophosaurus wetherilli* (Dinosauria, Theropoda) osteology and comparisons. *Palaeontogr. Abt. A* **185**, 85–180 (1984).
25. Nopcsa, F. On the origin of flight in birds. *Proc. Zool. Soc. Lond.* **1923**, 463–477 (1923).
26. Thulborn, R. A. The avian relationships of *Archaeopteryx*, and the origin of birds. *Zool. J. Linn. Soc.* **82**, 119–158 (1984).
27. Gauthier, J. Saurischian monophyly and the origin of birds. *Mem. Calif. Acad. Sci.* **8**, 1–55 (1986).
28. Gatesy, S. M. Hind limb movements of the American alligator (*Alligator mississippiensis*) and postural grades. *J. Zool. (Lond.)* **224**, 577–588 (1991).
29. Tarsitano, S. Stance and gait in theropod dinosaurs. *Acta Palaeontol. Pol.* **28**, 251–264 (1983).
30. Gatesy, S. M. Caudofemoral musculature and the evolution of theropod locomotion. *Paleobiology* **16**, 170–186 (1990).

Acknowledgements. We thank L. B. Clemmensen, the Danish Polar Center and the Commission for Scientific Research in Greenland for support of field work, L. L. Meszoly for rendering Fig. 1, M. A. Norell and A. R. Davidson for access to *Coelophysis bauri* (AMNH 7226), Alias/Wavefront for software support, and J. O. Farlow, J. R. Hutchinson, M. G. Lockley, P. E. Olsen, K. Padian, D. Norman and R. A. Thulborn for critical advice. This work was supported by grants from the National Science Foundation, the Carlsberg Foundation, and the Putnam Expeditionary Fund of Harvard University.

Correspondence should be addressed to S.M.G. (e-mail: Stephen_Gatesy@Brown.edu).

Evolution of genetic mechanisms controlling petal development

Elena M. Kramer & Vivian F. Irish

Department of Molecular, Cellular and Developmental Biology, Yale University, PO Box 208104, New Haven, Connecticut 06520–8104, USA

Molecular genetic studies in *Arabidopsis thaliana* and other higher-eudicot flowering plants have led to the development of the ‘ABC’ model of the determination of organ identity in flowers, in which three classes of gene, *A*, *B* and *C*, are thought to work together to determine organ identity^{1,2}. According to this model, the *B*-class genes *APETALA3* (*AP3*) and *PISTILLATA* (*PI*) act to specify petal and stamen identity. Here we test whether the roles of these genes are conserved throughout the angiosperms by

analysing the expression of *AP3* and *PI* orthologues in the lower eudicot subclass Ranunculidae. We show that, although expression of these orthologues in the stamens is conserved, the expression patterns in the petals differ from those found in the higher eudicots. The differences between these expression patterns suggest that the function of *AP3* and *PI* homologues as *B*-class organ-identity genes is not rigidly conserved among all angiosperms. These observations have important implications for understanding the evolution of both angiosperm petals and the genetic mechanisms that control the identities of floral organs.

The ABC model postulates that sepal identity is established by the functions of *A* genes alone, petals by *A* and *B*, stamens by *B* and *C*, and carpels by *C* function alone^{1,2}. Orthologous genes from the three classes have been identified in *Arabidopsis* and *Antirrhinum*¹, and subsequently in a number of other species, including *Nicotiana tabacum* (tobacco), *Solanum tuberosum* (potato) and *Petunia hybrida* (reviewed in ref. 3). The expression patterns and, where examined, the function of these genes appear to be similar in all of the species analysed^{3,4}, leading to the suggestion that the ABC model is conserved throughout the angiosperms⁵. However, all of the species in which the ABC genes have been well analysed are members of a monophyletic clade known as the higher eudicots. Although this group contains the majority of angiosperm species, it does not represent the full range of the flowering plants.

To assess the extent of conservation of the ABC model, we have

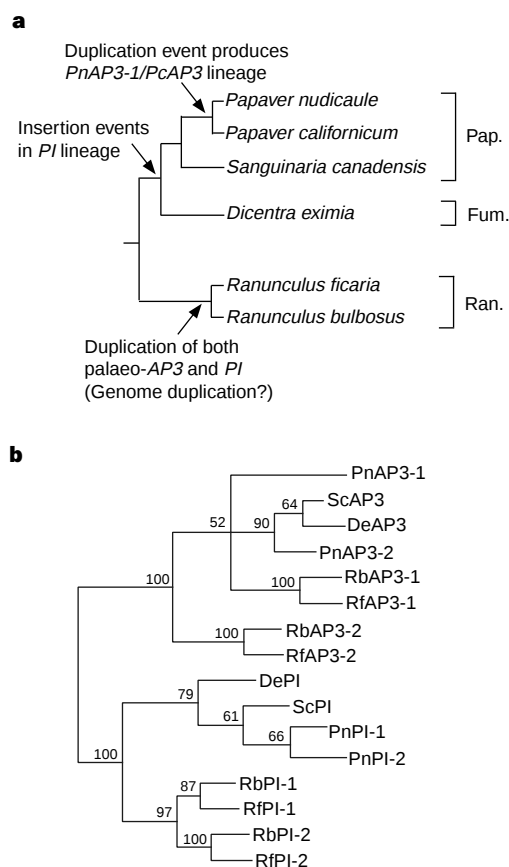


Figure 1 Gene and species phylogenies. **a**, Simplified phylogeny showing the relationships between Ranunculid species included in this study (based on ref. 22). Major events in the course of palaeo-*AP3* and *PI* gene evolution are shown. Pap.: Papaveraceae; Fum.: Fumariaceae; Ran.: Ranunculaceae. **b**, Gene tree of Ranunculid palaeo-*AP3* and *PI* lineage members based on parsimony analysis of nucleotide sequence. The numbers next to the nodes give bootstrap values from 1,000 replicates. A single tree of 1,839 steps was generated. Each cDNA was named using the first letter of the genus and species from which it was isolated followed by either *AP3* or *PI*, depending on its homology.

investigated the genetic mechanisms controlling floral development in angiosperm species outside the higher eudicots, specifically focusing on the *B*-class genes⁶. Two *B*-class gene lineages have been identified in the higher eudicots, the *Arabidopsis* representatives being the MADS-box genes *AP3* and *PI*^{7,8}. Phylogenetic analysis of the MADS-box gene family has shown that *AP3* and *PI* are representatives of closely related paralogous lineages⁹. The higher-eudicot *AP3* and *PI* orthologues are expressed in the presumptive petal and stamen primordia soon after the appearance of the sepal primordia, and expression is maintained in the petals and stamens as they develop^{3,4}. In *Arabidopsis* and *Antirrhinum*, continual expression of the *B*-class genes throughout these organ primordia, particularly the petals, is required for normal development^{10–12}. If the ABC model is indeed conserved throughout the angiosperms, *AP3* and *PI* homologues from species outside the higher eudicots should display the same pattern of expression. To test this prediction, we have characterized the expression patterns of *AP3* and *PI* orthologues in several lower-eudicot species.

We have cloned *AP3* and *PI* homologues from five species of the lower-eudicot subclass Ranunculidae (Fig. 1a). Two species are from the family Papaveraceae, *Papaver nudicaule* (Iceland poppy) and *Sanguinaria canadensis* (bloodroot), one is from its sister family Fumariaceae, *Dicentra eximia* (fringed bleeding-heart), and two are from the family Ranunculaceae, *Ranunculus bulbosus* (buttercup) and *R. ficaria* (lesser celandine). We have made an exhaustive effort to identify all of the florally expressed *AP3* and *PI* orthologues in

these species (at least 50 clones have been analysed for each species⁶).

A phylogeny of the *AP3* and *PI* orthologues from these five species was constructed (Fig. 1b). A single main *AP3* lineage, which we call the palaeo-*AP3* lineage, is present throughout the basal angiosperms and lower eudicots⁶. A duplication of the ancestral palaeo-*AP3* gene at the base of the higher eudicots appears to have given rise to two paralogous lineages: the eu-*AP3* lineage, including *AP3*, and the *TM6* lineage. The Ranunculid *AP3* homologues are all members of the palaeo-*AP3* lineage, and display all of the diagnostic palaeo-*AP3* sequence characteristics⁶. Within the Ranunculaceae, the grouping of the *AP3* and *PI* homologues indicates that a genome duplication, rather than two independent single-gene duplications, occurred before the divergence of the two *Ranunculus* species (Fig. 1a). The *P. nudicaule* gene *PnAP3-1* is quite divergent in sequence from *PnAP3-2* and the other Ranunculid palaeo-*AP3* representatives⁶. We have identified an orthologue of this gene in *Papaver californicum*⁶, but additional members of the lineage defined by *PnAP3-1* and *PcAP3* have not been detected by polymerase chain reaction (PCR) or Southern analysis in any of the Ranunculid species in this study (data not shown), suggesting that this lineage may have originated relatively recently within the Papaveraceae (Fig. 1a). The predicted product of the *S. canadensis* gene *ScPI* shares the distinctive carboxy-terminal insertions⁶ that appear to be synapomorphic for *PI* lineage in the Papaverales (Fig. 1a; *PnPI-2* encodes a truncated gene product⁶).

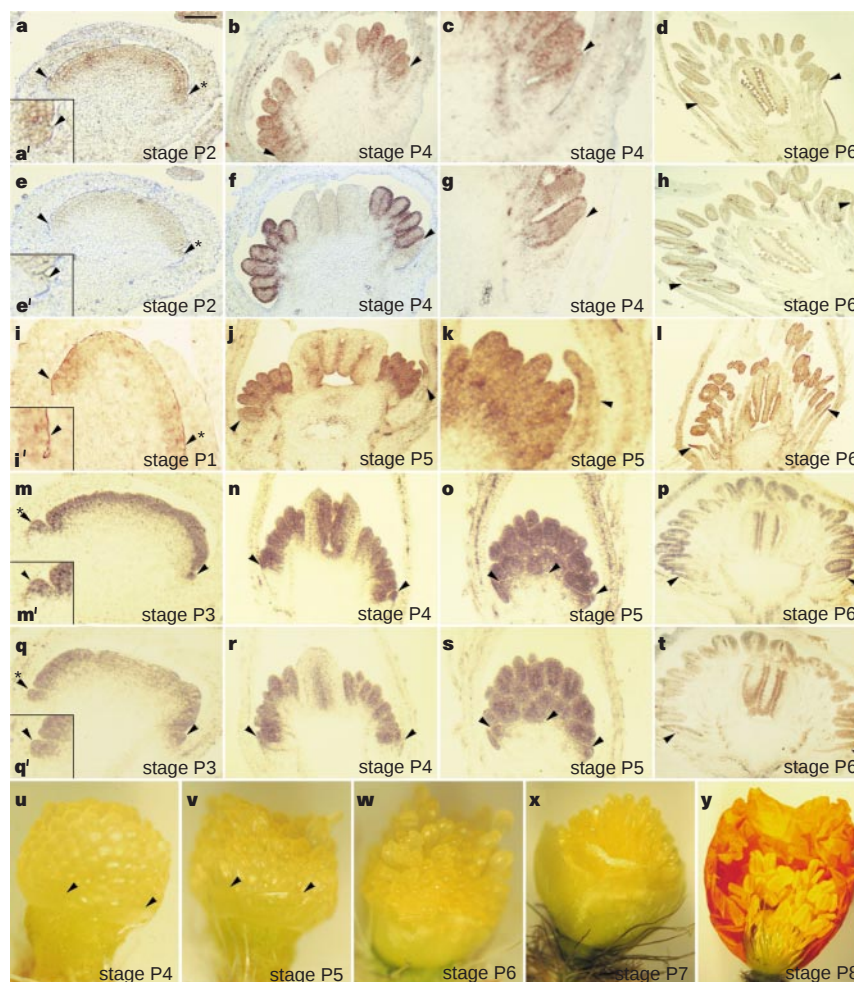


Figure 2 RNA and antibody localization in the developing *P. nudicaule* flower bud. Arrows indicate petal primordia. **a**–**d**, *PnAP3-1* expression. **e**–**h**, *PnAP3-2* expression. **i**–**l**, *PnPI-1* expression. **m**–**p**, anti-*AP3* staining. **q**–**t**, anti-*PI* staining. **u**–**y**, Stages P4–P8 of developing *P. nudicaule* flower. Bud stages are indicated in each panel. Both

labelled sense-RNA probes and secondary antibodies alone showed no specific staining (data not shown). Scale bar in **a** represents: **a**–**d**, **e**–**h**, **i**–**l**, **c**, **g**, 200 μ m; **a**, **e**, **o**, **s**, 400 μ m; **b**, **f**, **m**, **q**, 420 μ m; **d**, **l**, 925 μ m; **h**, 830 μ m; **i**, 380 μ m; **j**, 520 μ m; **k**, 180 μ m; **m**–**q**, 280 μ m; **n**, **r**, 650 μ m; **p**, **t**, 785 μ m; **u**, 700 μ m; **v**, 720 μ m; **w**, 1.1 mm; **x**, 2.75 mm; **y**, 8 mm.

The flowers of *P. nudicaule* are terminal, with two free sepals, four petals, numerous stamens and six fused carpels (the gynoecium). The sepal primordia are initiated first, followed by the petal primordia which are first detectable when the sepals have almost enclosed the floral dome (stage P1; Fig. 2i). The initiation of stamen primordia begins in the outermost regions of the apex and proceeds towards the centre (stage P2; Fig. 2a, e). The gynoecium is the last structure to initiate (stage P3; Fig. 2m, q). Once the stamen primordia have been formed, they develop rapidly (stages P4–5). In contrast, the petals grow slowly until the stamens have begun to mature, at which time they expand rapidly (stages P6–8; Fig. 2w–y).

The expression patterns of the *P. nudicaule* genes *PnAP3-1*, *PnAP3-2* and *PnPI-1* were characterized by *in situ* hybridization. During stages P1 and P2, both *PnPI-1* and *PnAP3-1* are expressed in the petal primordia (Fig. 2a, i). *PnAP3-2* is expressed at low levels at this stage (Fig. 2e). Once the stamens have begun to initiate, it is clear that *PnAP3-2* is not expressed in the petal primordia (Fig. 2f, g). The expression of *PnAP3-1* and *PnPI-1* in the petal diminishes when the stamen primordia become apparent (stage P4). Loss of expression begins in the abaxial base of the petal, then proceeds inwards and towards the tip (Fig. 2b, c, j, k, and data not shown). *PnAP3-1* and *PnPI-1* transcripts continue to be restricted to the extreme edges of the petals throughout their development (Fig. 2d, l). *PnAP3-2* expression turns on in the petals later in development, also only along the edges (Fig. 2h). Northern blot analyses have indicated that *PnPI-2* is expressed at very low levels (data not shown).

Expression of *PnAP3-1*, *PnAP3-2* and *PnPI-1* is relatively uniform and constant throughout the development of the stamens, as is seen in higher eudicots. Transcripts are first observed in the region giving rise to the stamen primordia, and expression is maintained throughout these organs until the locules begin to form, after which expression becomes localized to the sporangial tissue. *PnAP3-1*, *PnAP3-2* and *PnPI-1* are also expressed throughout the developing ovules until the initiation of integuments, when expression declines (Fig. 2b, j).

We examined the distribution of the AP3 and PI proteins in *P. nudicaule* using polyclonal antibodies raised against *Arabidopsis* AP3 and PI (P. Jenik and V.F.I., manuscript in preparation). This technique should detect all AP3- or PI-like proteins present in the tissue, rather than being gene-specific. In *P. nudicaule*, the anti-AP3 and anti-PI staining patterns (Fig. 2m–t) correspond exactly to the sum of the RNA expression patterns of their respective genes, further suggesting that we have cloned all of the AP3 and PI representatives in this species. The expression of these proteins in the very tips of the petal primordia is particularly striking (arrows in Fig. 2n–p, r–t). Glancing sections through early stage P4 buds reveal strongly staining stamens juxtaposed against the largely unstained petal primordia. Staining is only visible at the extreme edges of the petals (arrows in Fig. 2o, s).

In *D. eximia*, each terminal meristem produces one or two lateral meristems in the axil of a bract (stage D0; Fig. 3a, d), and then goes on to differentiate into a floral meristem. The lateral meristems reiterate this pattern. The two sepals are in anterior–posterior (medial) positions and do not tightly enclose the bud. The four petals show partial fusion and are arranged in two morphologically distinct, opposite pairs. It is the petals, rather than the sepals, that grow quickly early in floral development (stages D1 in Fig. 3b, e, and D2 in Fig. 3c, f) to enclose the reproductive organs. The six stamen primordia arise separately but become adherent just below the anthers to form two lateral bundles¹³ (stages D3–5 in Fig. 3g–l). The gynoecium arises as a ring primordium in the centre of the floral meristem.

Antibodies raised against AP3 and PI were used to characterize protein distribution in *D. eximia* flowers. During stage D0, both the terminal floral meristem and the lateral meristems express DeAP3 and DePI (Fig. 3a, d; also lm in Fig. 3c). This broad expression

pattern has been observed with floral organ and meristem identity genes in other species with terminal inflorescences^{14,15}. Expression of both proteins quickly becomes limited to the stamen primordia and the tips of the petal primordia (stage D1 in Fig. 3b, e). Some staining is also observed in the carpels (Fig. 3b, c, e, f). The petal primordia lose expression of both proteins during stage D2, first in the outer lateral petals, then in the inner medial petals (Fig. 3c, f, and data not shown). During stages D3 and D4, the proteins become concentrated in the stamens (especially the locules) and the placental region of the carpel, marking the nascent ovules (Fig. 3g, h, j, k). The distribution of anti-PI staining appears to be more diffuse than that of anti-AP3 during these stages. Also, anti-PI staining is often visible at the base of the petals, particularly in the epidermis and vasculature (Fig. 3j, k). As the bud approaches maturity (stage D5), expression of both proteins is seen in the developing ovules, the stamens and the base of the petals (Fig. 3i, l). The anti-PI signal

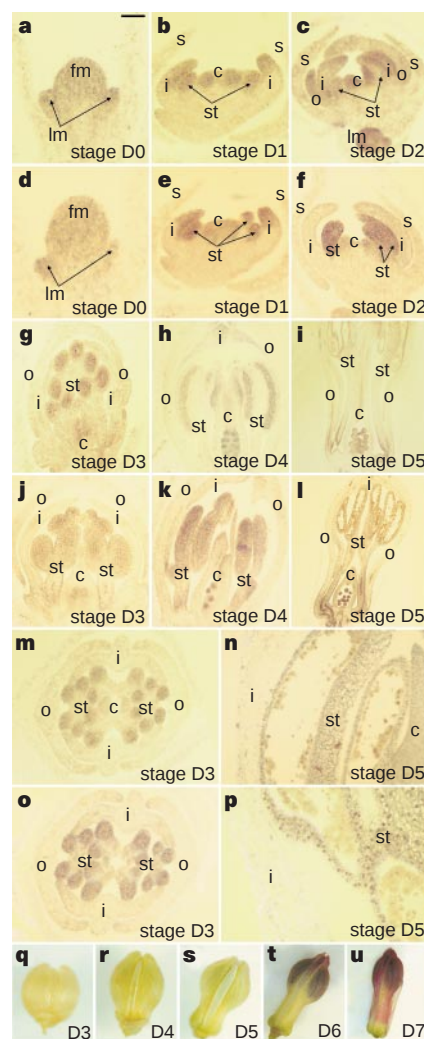


Figure 3 Antibody localization in the developing *D. eximia* flower bud. **a–c, g–i, m, n**, AP3 antibody. **d–f, j–l, o, p**, PI antibody. **a, d**, Frontal longitudinal sections of a stage D0 floral meristem (fm) with two lateral meristems (lm). **b, e**, Sagittal longitudinal sections of stage D1 buds. **c, f**, Sagittal longitudinal sections of stage D2 buds. **g, j**, Sagittal longitudinal sections of stage D3 buds. **h, k**, Frontal longitudinal sections of stage D4 buds. **i, l**, Frontal longitudinal sections of stage D5 buds. **m, o**, Transverse sections through D3 buds. **n, p**, Magnification of stamen and petal tissue in stage D5 bud. **q–u**, Stages D3–7 of developing *D. eximia* flower. fm, floral meristem; lm, lateral meristem; s, sepal; o, outer lateral petals; i, inner medial petals; st, stamen; c, carpel. Scale bar represents: **a, d**, 100 μ m; **b, e, n**, 115 μ m; **c, f**, 175 μ m; **g, j**, 380 μ m; **h, k**, 475 μ m; **i**, 715 μ m; **l**, 950 μ m; **m, o**, 250 μ m; **p**, 50 μ m; **q**, 700 μ m; **r**, 1.1 mm; **s**, 1.75 mm; **t**, 5 mm; **u**, 1 cm.

appears to be more intense than that of anti-AP3 in late stage buds. The sharp contrast between the absence of staining in the petals and strong signal in the stamens can be seen in transverse sections of stage-D3 bud (Fig. 3m, o) and in stage-D5 stamen and petal tissue (Fig. 3n, p). Northern blots of *DeAP3* and *DePI* transcripts in stages D3 onwards (Fig. 3q–u) are consistent with these results (data not shown).

Northern blots were used to analyse the expression patterns of all of the *AP3* and *PI* representatives in *R. bulbosus* and *R. ficaria* floral buds (Fig. 4a, b). The petals were separated into developmental stages roughly corresponding to those of *Papaver* and defined as follows: R4, petal primordia of less than a quarter of the height of the stamens; R5, petals ranging in size from a quarter to half the height of the stamens; R6, petals ranging in size from a quarter to half the height of the stamens; R6, petal size ranging from half to the full height of the stamens; R7, petals of the same height as the stamens or longer; and R8, fully mature petals.

The expression patterns of *AP3* and *PI* representatives in the *Ranunculus* species vary considerably (Fig. 4a, b). In *R. bulbosus*, all four genes are expressed in a very similar pattern: expression is low

in stage-R4 petal primordia and increases during later stages. These genes also show high levels of expression in the stamens. *R. ficaria* displays a considerably different pattern: *RfAP3-1* has extremely low petal expression until stage R8, whereas *RfAP3-2* is expressed at low levels earlier in petal development, during stages R4–R6. These patterns differ from each other and from those of their *R. bulbosus* orthologues. The expression of the *RfPI-1* and *RfPI-2* genes is comparatively low in both petal and stamen primordia.

Among the *Ranunculid* species we examined, neither uniform nor constant expression throughout petal development was observed for any of the *AP3* or *PI* homologues. This is in contrast to the higher eudicots where, with some small variation, all of the *AP3* and *PI* orthologues analysed so far are expressed throughout the developing petals at all stages, as well as in the stamen primordia (reviewed in ref. 3). Moreover, this ubiquitous petal expression pattern is necessary for normal petal development in both *Arabidopsis* and *Antirrhinum*^{10–12}. We have therefore reconsidered the idea that the ABC programme as defined in the higher eudicots is rigidly conserved across all the angiosperms. An explanation for these results is based on classical morphological studies which suggest that the petals have been derived independently many times from either stamens or sterile bracts, depending on the angiosperm lineage¹⁶. These independent evolutionary events would logically have involved the recruitment of genes to new roles, producing novel expression patterns in the petals. It remains likely that the 'B function' of the higher eudicot *AP3* and *PI* orthologues, that is, their requirement for specifying both petal and stamen identity, is conserved among the higher eudicots. The *Ranunculid* palaeo-*AP3* and *PI* representatives do not, however, appear to function in the same manner to specify petal identity. Their roles in stamen development may be conserved, reflecting a single evolution of stamens, but the diverse expression patterns we have observed in petals may indicate one or more independent derivation events in which the *AP3* and *PI* representatives were recruited to new tissue- or cell-type-specific roles in the petals. The *Ranunculidae* are considered to have staminally derived petals which are thought to have arisen from many independent evolutionary events^{17,18}. Our results indicate that, in the *Ranunculids*, other genes may have been recruited to perform the role of specifying petal identity, either alone or in combination with the *AP3* and *PI* orthologues. Functional analyses of these genes will allow us to characterize their roles in flower development more fully. □

Methods

Phylogenetic analysis. Twelve genes analysed here have been previously reported⁶. Additional B-group orthologues from *S. canadensis*, *R. bulbosus* and *R. ficaria* were cloned as described⁶. Alignment of nucleotide sequences and parsimony analysis were done as described⁶. The sequences reported in this Letter have been deposited with GenBank.

Southern and northern blot analysis. Blotting was done as described¹⁹, except that northern blots were hybridized at 43 °C in 50% formamide, 3 × SSC, 0.5% SDS, 0.1 mg ml⁻¹ single-stranded herring sperm DNA, 5 × Derrhardt's solution and 25 mM EDTA, pH 8, and washed at 50 °C in 1 × SSC, 0.5% SDS. All random-primed DNA probes were made from gene-specific templates which did not include the MADS or K domains.

In situ expression analysis in *Papaver nudicaule*. This procedure has been described²⁰. Digoxigenin-labelled RNA probes for *PnAP3-1*, *PnAP3-2* and *PnPI-1* were prepared from linearized templates cloned into pCR2.1 plasmids (Invitrogen). RNA probes were gene-specific and did not include the MADS or K domains.

Immunohistochemistry. Polyclonal antibodies were generated against amino acids 79–232 of *AP3* and 73–208 of *PI* (P. Jenik and V.F.I., manuscript in preparation). Immunohistochemistry using *AP3*- and *PI*-specific antibodies in excess was as described²¹.

Received 25 January; accepted 9 March 1999.

- Coen, E. S. & Meyerowitz, E. M. The war of the whorls: genetic interactions controlling flower development. *Nature* **353**, 31–37 (1991).

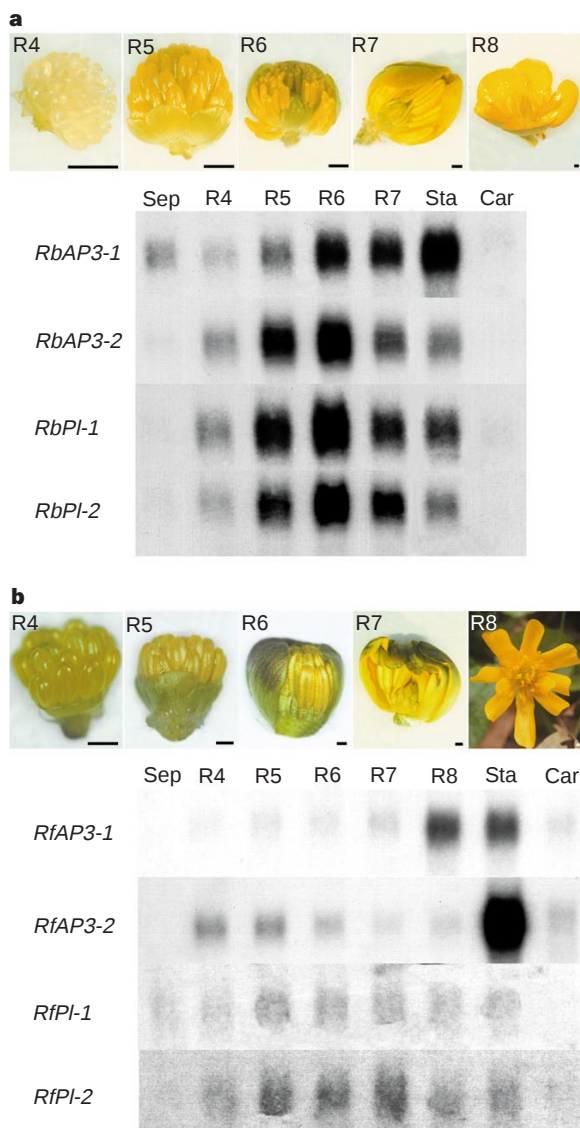


Figure 4 Northern blot analysis of *Ranunculus* species. **a**, stages R4–R8 for *R. bulbosus*, with northern blot results for *RbAP3-1*, *RbAP3-2*, *RbPI-1*, *RbPI-2*. Scale bars, 1 mm. **b**, Stages R4–R8 for *R. ficaria* with northern blot results for *RfAP3-1*, *RfAP3-2*, *RfPI-1*, *RfPI-2*. Scale bars, 1 mm. All northern blot analyses were carried out with 10 µg RNA per lane. Lane designations: Sep, sepals; R4–R8, petal primordia collected at given stages (see text); Sta, stamens; Car, carpels.

2. Meyerowitz, E. M. *et al.* A genetic and molecular model for floral development in *Arabidopsis thaliana*. *Development* **1**, 157–167 (1991).
3. Irish, V. F. & Kramer, E. M. in *Advances in Botanical Research* 197–230 (Academic, San Diego, 1998).
4. Theissen, G. & Saedler, H. MADS-box genes in plant ontogeny and phylogeny Haeckel's 'biogenetic law' revisited. *Curr. Opin. Gen. Dev.* **5**, 628–639 (1995).
5. Bowman, J. L. Evolutionary conservation of angiosperm flower development at the molecular and genetic levels. *J. Biosci.* **22**, 515–527 (1997).
6. Kramer, E. M., Dorit, R. L. & Irish, V. F. Molecular evolution of genes controlling petal and stamen development: Duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. *Genetics* **149**, 765–783 (1998).
7. Jack, T., Brockman, L. L. & Meyerowitz, E. M. The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* **68**, 683–697 (1992).
8. Goto, K. & Meyerowitz, E. M. Function and regulation of the *Arabidopsis* floral homeotic gene *PISTILLATA*. *Genes Dev.* **8**, 1548–1560 (1994).
9. Purugganan, M. D., Rounsley, S. D., Schmidt, R. J. & Yanofsky, M. F. Molecular evolution of flower development: Diversification of the plant MADS-box regulatory gene family. *Genetics* **140**, 345–356 (1995).
10. Bowman, J. L., Smyth, D. R. & Meyerowitz, E. M. Genes directing flower development in *Arabidopsis*. *Plant Cell* **1**, 37–52 (1989).
11. Zachgo, S. *et al.* Functional analysis of the *Antirrhinum* floral homeotic *Deficiens* gene *in vivo* and *in vitro* by using a temperature-sensitive mutant. *Development* **121**, 2861–2875 (1995).
12. Sommer, H. *et al.* Properties of *deficiens*, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*. *Development* **1** (suppl.), 169–175 (1991).
13. Ronse Decraene, L. P. & Smets, E. F. An updated interpretation of the androecium of the Fumariaceae. *Can. J. Bot.* **70**, 1765–1776 (1992).
14. Kelly, A. J., Bonnländer, M. B. & Meeks-Wagner, D. R. NFL, the tobacco homolog of *floricaula* and *leafy*, is transcriptionally expressed in both vegetative and floral meristems. *Plant Cell* **7**, 225–234 (1995).
15. Hardenack, S., Ye, D., Saedler, H. & Grant, S. Comparison of MADS box gene expression in developing male and female flowers of the dioecious plant white campion. *Plant Cell* **6**, 1775–1787 (1994).
16. Takhtajan, A. *Evolutionary Trends in Flowering Plants* (Columbia Univ. Press, New York, 1991).
17. Drinnan, A. N., Crane, P. R. & Hoot, S. B. in *Early Evolution of Flowers* (eds Endress, P. K. & Friis, E. M.) 93–122 (Springer, New York, 1994).
18. Kosuge, K. Petal evolution in Ranunculaceae. *Plant Syst. Evol.* **8** (suppl.), 185–191 (1994).
19. Carr, S. & Irish, V. Floral homeotic gene expression defines developmental arrest stages in *Brassica oleracea* L. vars. *botrytis* and *italica*. *Planta* **201**, 179–188 (1997).
20. Lincoln, C., Long, J., Yamaguchi, J., Serikawa, K. & Hake, S. A knotted1-like homeobox gene in *Arabidopsis* is expressed in the vegetative meristem and dramatically alters leaf morphology when overexpressed in transgenic plants. *Plant Cell* **6**, 1859–1876 (1994).
21. Smith, L., Greene, B., Veit, B. & Hake, S. A dominant mutation in the maize homeobox gene, *Knotted-1*, causes its ectopic expression in leaf cells with altered fates. *Development* **116**, 21–30 (1992).
22. Chase, M. W. *et al.* Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Ann. Missouri Bot. Gard.* **80**, 528–580 (1993).

Acknowledgements. We thank P. Jenik for antibody preparation, and I. Dawson, P. Jenik, T. Hill, C. Juarez, R. Lamb and Q. Tan for comments on the manuscript. This work was supported by grants from the NSF and USDA to V.F.I.

Correspondence and requests for materials should be addressed to V.F.I. (e-mail: vivian.irish@yale.edu).

The role of the anterior prefrontal cortex in human cognition

Etienne Koechlin, Gianpaolo Basso, Pietro Pietrini, Seth Panzer & Jordan Grafman

Cognitive Neuroscience Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892-1440, USA

Complex problem-solving and planning involve the most anterior part of the frontal lobes including the fronto-polar prefrontal cortex (FPPC)^{1–6}, which is especially well developed in humans compared with other primates^{7,8}. The specific role of this region in human cognition, however, is poorly understood. Here we show, using functional magnetic resonance imaging, that bilateral regions in the FPPC alone are selectively activated when subjects have to keep in mind a main goal while performing concurrent (sub)goals. Neither keeping in mind a goal over time (working memory) nor successively allocating attentional resources between alternative goals (dual-task performance) could by themselves activate these regions. Our results indicate that the FPPC selectively mediates the human ability to hold in mind goals while exploring and processing secondary goals, a process generally required in planning and reasoning.

This process of integrating working memory with attentional resource allocation is referred to as branching. In everyday life, branching is frequently required: for example, if a person is inter-

Table 1 Factorial design of the experiment

Attentional resource allocation	Working memory	
	No delayed response	Delayed response
Single-task performance	Control	Delay
Dual-task performance	Dual task	Branching

rupted with a question while reading. As in dual-task performance, branching successively allocates processing resources between concurrent tasks (such as listening and reading). As in delayed-response performance, branching keeps relevant information in working memory to allow a return to the main task after completing a secondary task (in our example, remembering and returning to where you left off reading). Knowing that problem-solving and planning involve the FPPC^{1–6} and generally involve branching, whenever goal-tree sequences are processed we proposed that only fronto-polar regions would be selectively engaged by the integration of working memory and attentional-resource allocation.

Using functional magnetic resonance imaging (fMRI), we employed a 2 × 2 factorial design crossing delayed-response and dual-task performance (Table 1), so that subjects performed an on-line letter-matching task in four conditions including a control, a delay, a dual-task and a branching condition (see Figs 1 and 2). Accordingly, specific branching activations were specified as an interaction between delayed-response and dual-task performance and were expected to be found only in fronto-polar regions.

The fMRI results confirm our prediction (Table 2 and Figs 3–5). First, we tested for regions selectively involved in either dual- or delayed-task performance (see Methods). The main effect of dual-task performance was found to involve, selectively and bilaterally, the posterior dorsolateral prefrontal cortex (BA (Brodmann's area) 9, middle frontal gyrus, close to the pre-central gyrus and BA 8) and the lateral parietal cortex (inferior parietal lobule, BA 40). In accordance with previous studies, this confirms that this network of brain areas is involved in allocating attentional resources between successively alternating tasks or stimuli^{9,10}. In contrast, the main effect of delayed performance involved no specific brain region even

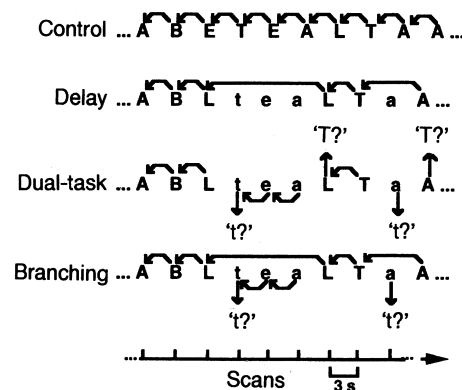


Figure 1 Behavioural tasks. Stimuli were pseudorandom sequences of upper- or lower-case letters from the word 'tablet'. Control condition: subjects had to decide whether two successively presented letters were also in immediate succession in the word 'tablet' (only upper-case letters were presented). Delay condition: subjects had to ignore lower-case letters which were used to occasionally delay the response required by the control condition. Dual task condition: subjects had to respond as in the control condition for both upper- and lower-case letter series with one exception. Subjects had to decide whether every first letter indicating a case change was the letter T (or t). Branching condition: subjects had to respond to upper-case letters exactly as in the delay condition and to lower-case letters exactly as in the dual-task condition.

Table 2 Coordinates and Z-score for maxima of activations

Regions Brodman's area*	Talairach coordinates†	Statistical effects (Z-value)‡						
		BR-CR	DE-CT	DT-CT	BR-DE	BR-DT	DE-DT	DT-DE
Interaction								
Right, BA 10, GFp	36, 66, 21	8.2	n.s.	n.s.	6.8	6.8	n.s.	n.s.
Left, BA 10, GFp	-36, 57, 9	8.1	n.s.	n.s.	6.2	7.3	n.s.	n.s.
Additivity								
Right, BA 10, GFs	30, 75, 12	8.5	6.8	7.6	7.0	5.2	n.s.	n.s.
Right, BA9/46, GFm	54, 36, 30	8.5	6.5	7.9	7.1	3.9	n.s.	n.s.\$
Right BA 7, Pcu	6, -60, 48	8.3	6.1	7.5	6.1	3.6	n.s.	n.s.
Main effects (dual task)								
Right, BA 9/8, GFm	45, 21, 36	8.3	5.3	8.1	6.8	n.s.	n.s.	5.6
Right, BA 40, LPi	48, -39, 48	7.7	n.s.	8.0	6.5	n.s.	n.s.	7.3
Left, BA 9/8, GFm	-45, 27, 36	7.9	n.s.	7.3	7.2	n.s.	n.s.	5.6
Left, BA 7, LPs	-42, -51, 54	8.3	n.s.\$	8.2	7.7	n.s.	n.s.	7.3

Data relate to maxima of activations shown in Fig. 3.

* BA, Brodmann's area; GF, frontal gyrus; LP, parietal lobule; l, inferior; m, middle; s, superior; p, polar; PCu, precuneus.

† Talairach coordinates of maxima in contrast branching minus control conditions.

‡ BR, branching; DE, delay; DT, dual-task; CT, control. n.s. not significant ($P > 0.05$, corrected).

§ $P > 0.5$, corrected.

at low statistical thresholds. This indicates that the duration of delay during which information is maintained on-line in this task is not in itself a main factor triggering the activation of specific cortical regions.

Our main aim was to identify brain regions involved in both delayed-response and dual-task performance. Those regions fall into one of two categories: regions that exhibit additive effects, where the evoked response is simply explained by the activation due to utilizing working memory or attention allocation independently; and regions that show an interaction, where there is a super-additive component that is attributable to the cognitive integration of working memory and attentional allocation (in our framework, a branching process).

Additive activations were obtained by selecting the regions engaged in all experimental conditions compared with baseline (see Methods). No interaction was observed *post hoc* in those regions ($P > 0.05$), namely the right anterior dorsolateral prefrontal cortex (BA 9/46, middle frontal gyrus), the right superior frontal gyrus (BA 10) and the right precuneus (BA 7) (Figs 3 and 5). When compared to the control condition, the MR signal change recorded

in these brain regions during the branching condition co-varied with the sum of the signal changes in the delay and dual-task conditions. This cortical network is involved in episodic memory tasks, becoming activated when the circumstances in which events occurred are retrieved^{11–16}. In accordance with these results, all but the control conditions in our experiment share the common feature that subject performance depends upon processing and retrieving the context in which a stimulus occurs (either its lower/upper case status on a given trial or the task rules on a given experimental block). Performance in the branching condition required subjects to combine the previously independent contextual information used in the delay and dual-task conditions, which may explain the additivity effect found in these regions.

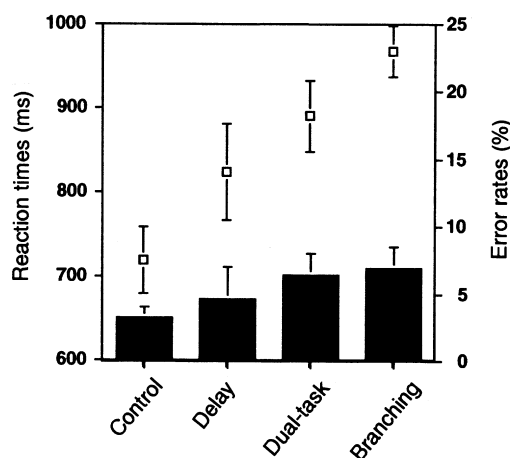


Figure 2 Behavioural performance. Symbols: increases in reaction times (mean $\pm$ s.e. in ms) across conditions ($F(3, 15) = 14.6$, $P < 0.0001$), confirming that additional processes and regions are engaged successively in the control, delay, dual-task and branching conditions. Bars: error rates (mean $\pm$ s.e. in percentage) remained very similar across conditions ($F(3, 15) = 1.2$, $P > 0.34$) with virtually no difference between the dual-task and branching conditions.

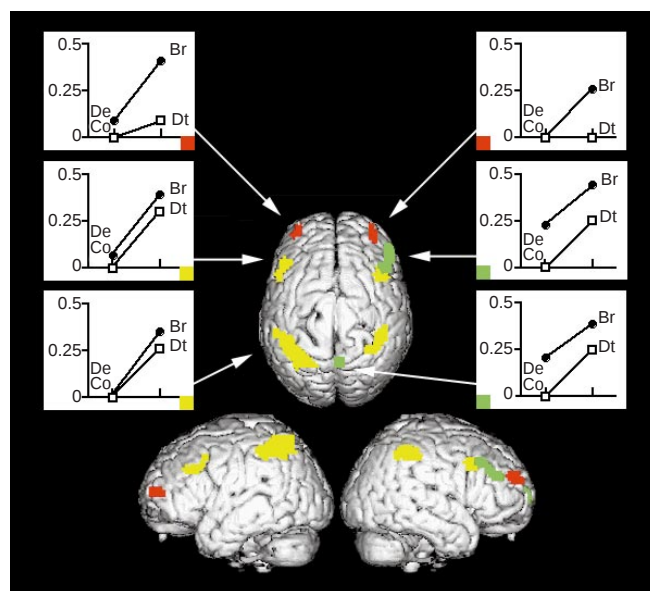


Figure 3 Topography of brain regions with distinct activation profiles. Yellow, main effects of dual-task performance. Green, additivity of the dual-task and delayed-response performance effects. Red, Interaction of dual-task and delayed-response effects or branching-specific activations. See Methods for details and Table 2 for coordinates of activation foci. Inserts, data points are the mean signal changes (vertical axis, percentage) in the delay (De), dual-task (Dt) and branching (Br) conditions (measured in the stationary state, that is, in the second half-block) relative to the adjusted signal mean in the control condition (Co).

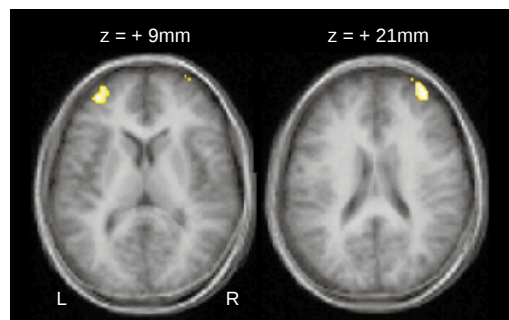


Figure 4 Branching-specific activations (Z -maps thresholded at $Z > 4.5$, $P < 0.05$, corrected) superimposed on anatomical axial slices averaged across subjects (Talairach coordinates $Z = 9$ and 21 mm). See Fig. 3 for details.

Super-additive, or branching-specific, activation was obtained by selecting regions which were engaged in the branching condition compared with baseline and, in addition, showed a significant interaction between working memory and attentional allocation factors (see Methods). As predicted, this revealed only two regions, located symmetrically in the left and right dorsal FPPC (BA 10, fronto-polar gyrus; Figs 3 and 4). In both regions, the MR signal increased significantly in the branching condition, but was quantitatively similar in all other conditions (Fig. 5). As neither the delayed nor dual-task performance alone significantly activated these regions, these results indicate that these fronto-polar regions are crucial only when branching processes are engaged. The activated foci differed slightly in the left and right fronto-polar regions, which may result from variations in the anatomy or the functional mapping between the two frontal lobes¹⁷.

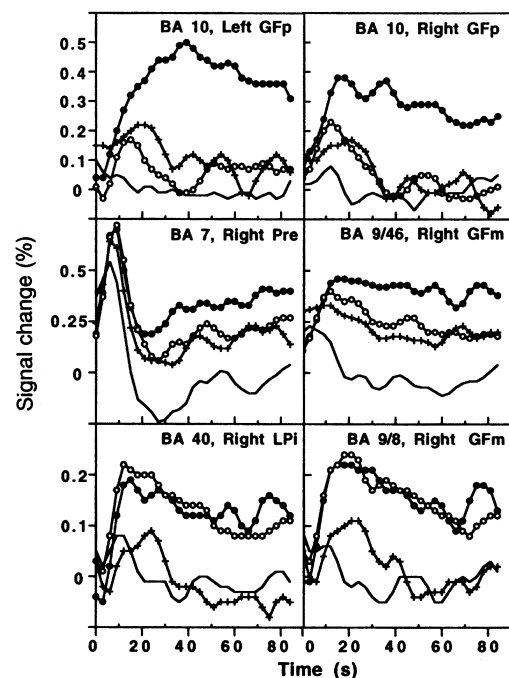


Figure 5 Dynamics of activation profiles. Plots are related to the regions shown in Fig. 3 and Table 2; x-axis, time measured from experimental block onsets. Stimulus onsets occur every 3 s, starting from time 3. Data points are relative signal changes averaged in every region during branching (filled circle), delay (cross), dual-task (circle) and control (no symbol) conditions. Signal changes (in per cent) are measured from the adjusted signal mean in the control condition.

Behavioural data show that these fronto-polar activations were not related to variations in mental effort alone. If that had been the case, a gradual increase in MR signal, associated with the gradual increase in difficulty of the tasks from the control to the delayed-task, dual-task and branching conditions (Fig. 2), would have been observed¹⁸. Instead, no difference in brain activation was found in these regions between conditions with different behavioural performances (for example, the control and dual-task conditions), whereas significant activation differences were observed between the dual-task and branching conditions despite very similar behavioural performance. Moreover, six more normal right-handed subjects (three males and three females, aged 20–28) performed the dual-task condition under more difficult constraints using degraded letters. Although subjects could still identify all letters, the task was significantly more difficult than the branching condition. Response times increased by 112 ms from the regular to the difficult dual-task condition ($P < 0.02$), and mean accuracy fell from 96.2 to 87.7%. As expected, increasing difficulty resulted in increased frontal activations in the posterior dorsolateral prefrontal cortex and the anterior cingulate cortex, but not in the FPPC ($Z < 1.0$, $P > 0.15$) in contrast to the results predicted by a mental effort interpretation of our data.

In summary, distinct dorsal fronto-polar regions were selectively engaged in branching processes when compared with other executive processes including working memory^{19–22}, attentional resource allocation^{9,10} and processing episodic information^{11–16}. This finding may clarify why the FPPC was found to be engaged in complex problem-solving^{1–6} and, incidentally, in working-memory tasks performed in dual-task contexts^{2,3,23,24}. These branching-specific regions may have a key role in processing goal-tree sequences, which frequently requires the temporary interruption of a current plan to achieve subgoals, or to respond to new environmental demands or intrusive thoughts, and may help to mediate a range of human behaviours including planning and reasoning. □

Methods

Behavioural protocol. Subjects responded to visually presented letters (500 ms duration, 3,000 ms stimulus-onset-asynchrony (SOA)) by pressing response buttons with their right (match) or left (no match) hand, respectively. Matching proportions were maintained between 40 and 43% of trials in each condition. In all but the control conditions lower-case letters were pseudorandomly presented in 64% of trials and the mean SOA between two successive upper-case letters was strictly maintained at 6.3 s. In the delay condition, subjects were asked to ignore lower-case letters by always pressing the no-match button (see Fig. 1 legend for details). The task was administered in six scanning runs using the Expe software package²⁵. Each condition was included once in each run as a block of 28 trials. The resulting 24 blocks were pseudorandomly ordered so that each condition appears at all serial positions within a run and two conditions appeared once or twice in immediate succession to prevent confounding order effects. Subjects were given standard instructions to respond quickly and accurately. Behavioural and fMRI data were recorded simultaneously from six right-handed subjects (three females and three males, aged 20–28 years) who were trained and reached a threshold accuracy ($> 70\%$) in all conditions before and during scanning. Three additional subjects were excluded because they did not reach this threshold during scanning. Subjects provided written informed consent approved by the NIH.

Image acquisition and analysis. A standard 1.5 GE signa scanner whole-body and RF coil scanner were used to perform a high-resolution structural scan for each subject followed by six runs of 120 functional axial scans (TR 3 s, TE 40 ms, flip angle 90° ; FOV 24 cm, acquisition matrix 64×64 , number of slices 18, thickness 6 mm) synchronized with stimulus presentation. All fMRI data were processed using the SPM96 software package (<http://www.fil.ion.ucl.ac.uk/spm/>)²⁶ with modified memory-mapping procedures. Standard linear image realignment, linear normalization to the stereotaxic Talairach atlas (MNI template)²⁷, spatial (3D gaussian kernel: 10 mm) and temporal smoothing, and mean MR-signal normalization across scans were successively performed for each subject. Then, all subjects were pooled together and statistical

parametric maps (SPM) were computed from local MR signals using a linear multiple regression with conditions (modelled as two temporal basis functions) and runs as co-variables²⁶. Only regions formed by more than 12 contiguous active voxels ($P < 0.05$) were analysed. The main effect of dual-task (and delayed-response, respectively) performance was computed by selecting the regions which were co-jointly significantly activated in the dual-task (delay) and branching conditions compared with the control and delay (dual-task) conditions ($Z > 5.4$, $P < 0.0005$, corrected for multiple comparisons). The additive effect of delay and dual-task performance was observed in the regions co-jointly and significantly activated in all conditions compared with the baseline ($Z > 5.4$, $P < 0.0005$, corrected). No interaction between those factors ($P > 0.05$, uncorrected) was observed in those regions. Branching-specific activations were computed as the regions with significant activations in the branching condition compared with the control ($Z > 5.4$, $P < 0.0005$, corrected) and with a significant interaction between the delayed-response and dual-task factors (branching and control compared to delay and dual-task conditions; $P < 0.0005$, uncorrected). The same branching-specific activation was found *post hoc* by computing the voxels co-jointly activated in the branching condition compared separately with all other conditions (fronto-polar: $Z > 5.4$, $P < 0.0005$, corrected; elsewhere $P > 0.05$, corrected). Branching-specific activations were confirmed in five single-subject analyses. They were located bilaterally in the superior fronto-polar gyrus at the crossing of the middle frontal gyrus (two subjects) or in the middle fronto-polar gyrus (three subjects)²⁸. Moreover, bilateral fronto-polar activations were replicated in six additional normal right-handed subjects performing the branching and two additional control tasks.

Received 23 October 1998; accepted 12 March 1999.

- Grafman, J. in *Structure and Function of the Human Prefrontal Cortex* (eds Grafman, J., Holyoak, K. J. & Boller, F.) 337–368 (Annals of the New York Academy of Sciences, New York, 1995).
- Baker, S. C. et al. Neural system engaged by planning: A PET study of the Tower of London task. *Neuropsychologia* **34**, 515–526 (1996).
- Owen, A. M., Doyon, J., Petrides, M. & Evans, A. C. Planning and spatial working memory: a positron emission tomography study in humans. *Eur. J. Neurosci.* **8**, 353–364 (1996).
- Sirigu, A. et al. Planning and script analysis following prefrontal lobe lesions. *Ann. NY Acad. Sci.* **769**, 277–288 (1995).
- Spector, L. & Grafman, J. in *Handbook of Neuropsychology* (eds Boller, F. & Grafman, J.) 377–392 (Elsevier, Amsterdam, 1994).
- Wharton, C. & Grafman, J. Reasoning and the human brain. *Trends Cogn. Sci.* **2**, 54–59 (1998).
- Stuss, D. T. & Benson, D. F. *The Frontal Lobes* (Raven, New York, 1986).
- Fuster, J. M. *The Prefrontal Cortex. Anatomy, Physiology, and Neuropsychology of the Frontal Lobe* (Raven, New York, 1989).
- D'Esposito, M. et al. The neural basis of the central executive system of working memory. *Nature* **378**, 279–281 (1995).
- Corbetta, M. Frontoparietal cortical networks for directing attention and the eye to visual locations: identical, independent, or overlapping neural systems? *Proc. Natl Acad. Sci. USA* **95**, 831–838 (1998).
- Squire, L. R. et al. Activation of the hippocampus in normal humans: a functional anatomical study of memory. *Proc. Natl Acad. Sci. USA* **89**, 1837–1841 (1992).
- Buckner, R. L. et al. Functional anatomical studies of explicit and implicit memory retrieval tasks. *J. Neurosci.* **15**, 12–29 (1995).
- Shallice, T. et al. Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature* **368**, 633–635 (1994).
- Grady, C. L. et al. Age-related reductions in human recognition memory due to impaired encoding. *Science* **269**, 218–221 (1995).
- Tulving, E. et al. Neuroanatomical correlates of retrieval in episodic memory: auditory sentence recognition. *Proc. Natl Acad. Sci. USA* **91**, 2012–2015 (1994).
- Tulving, E. *Elements of Episodic Memory* (Oxford Science, Oxford, 1983).
- Galaburda, A. M., LeMay, M., Kemper, T. L. & Geschwind, N. Right-left asymmetries in the brain. *Science* **199**, 852–856 (1978).
- Furey, M. L. et al. Cholinergic stimulation alters performance and task-specific regional cerebral blood flow during working memory. *Proc. Natl Acad. Sci. USA* **94**, 6512–6516 (1997).
- Goldman-Rakic, P. C. in *Handbook of Physiology, The Nervous System* (eds Plum, F. & Mountcastle, V.) 373–417 (American Physiological Society, Bethesda, MD, 1987).
- Petrides, M. in *Handbook of Neuropsychology* (eds Boller, F. & Grafman, J.) 59–82 (Elsevier, Amsterdam, 1994).
- Jonides, J. et al. Spatial working memory in humans as revealed by PET. *Nature* **363**, 623–625 (1993).
- Courtney, S. M., Ungerleider, L. G., Keil, K. & Haxby, J. V. Transient and sustained activity in a distributed neural system for human working memory. *Nature* **386**, 608–611 (1997).
- MacLeod, A. K., Buckner, R. L., Miezin, F. M., Petersen, S. E. & Raichle, M. E. Right anterior prefrontal cortex activation during semantic monitoring and working memory. *NeuroImage* **7**, 41–48 (1998).
- Grafton, S. T., Hazeltine, E. & Ivry, R. Functional mapping of sequence learning in normal humans. *J. Cogn. Neurosci.* **7**, 497–510 (1995).
- Pallier, C., Dupoux, E. & Jeannin, X. EXPE: an expandable programming language for on-line psychological experiments. *Behav. Res. Methods Instrum. Comput.* **29**, 322–327 (1997).
- Friston, K. J., Frith, C. D., Liddle, P. F. & Frackowiak, R. S. J. Comparing functional (PET) images: The assessment of significant change. *J. Cereb. Blood Flow Metab.* **11**, 690–699 (1991).
- Talairach, J. & Tournoux, P. *Co-planar Stereotaxic Atlas of the Human Brain* (Thieme Medical, New York, 1988).
- Duvernoy, H. *The Human Brain. Surface, Three Dimensional Sectional Anatomy and MRI*. (Springer, Wien, 1991).

Correspondence and requests for materials should be addressed to J.G. (e-mail: jgr@box-j.nih.gov).

Synaptic calcium transients in single spines indicate that NMDA receptors are not saturated

Zachary F. Mainen, Roberto Malinow & Karel Svoboda

Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

At excitatory synapses in the central nervous system, the number of glutamate molecules released from a vesicle is much larger than the number of postsynaptic receptors. But does release of a single vesicle normally saturate these receptors? Answering this question is critical to understanding how the amplitude and variability of synaptic transmission are set and regulated. Here we describe the use of two-photon microscopy¹ to image transient increases in Ca^{2+} concentration mediated by NMDA (N-methyl-D-aspartate) receptors in single dendritic spines of CA1 pyramidal neurons in hippocampal slices. To test for NMDA-receptor saturation, we compared responses to stimulation with single and double pulses. We find that a single release event does not saturate spine NMDA receptors; a second release occurring 10 ms later produces ~80% more NMDA-receptor activation. The amplitude of spine NMDA-receptor-mediated $[\text{Ca}^{2+}]$ transients (and the synaptic plasticity which depends on this) may thus be sensitive to the number of quanta released by a burst of action potentials and to changes in the concentration profile of glutamate in the synaptic cleft.

The issue of receptor saturation has been controversial, partly owing to complexities in the interpretation of indirect experiments². A direct test for saturation is to examine the interaction between two synaptic currents occurring in rapid sequence^{3,4}. Receptor saturation during the first event will block the response to a second event that occurs before transmitter has had a chance to unbind from the receptors. Glutamate unbinding from NMDA receptors occurs slowly ($\tau > 100$ ms)^{5,6}. Although it is conceptually simple, this experiment is complicated by the probabilistic nature of transmitter release. Because only a fraction of stimulated synapses release transmitter, the first and second responses might not occur at the same synapses, as is required for the interaction test. A reliable way to detect failure and release of transmitter at a single synapse is therefore required.

We used two-photon-laser scanning microscopy (2PLSM)¹ to monitor synaptic activation by measuring $[\text{Ca}^{2+}]$ transients in postsynaptic spines⁷. Dendritic spines on CA1 pyramidal cells have a single postsynaptic density⁸ associated with a single apposing presynaptic active zone⁹. Ca^{2+} entering spines through NMDA-receptor (NMDA-R) channels is therefore a direct reflection of NMDA-R activation at single synapses. We filled neurons in brain slices by whole-cell recording with a high-affinity Ca^{2+} indicator and activated synapses by extracellular stimulation using a glass pipette placed close to an apical dendritic branch¹⁰. Synaptic $[\text{Ca}^{2+}]$ transients were imaged using either frame (16 Hz; Fig. 1) or line-scan (500 Hz; Fig. 2) mode. Frame mode allowed us to discriminate between anatomically distinct Ca^{2+} sources and thereby to isolate $[\text{Ca}^{2+}]$ signals arising from single spines (Fig. A in Supplementary Information).

To isolate NMDA-R-mediated Ca^{2+} influx, neurons were voltage-clamped at positive potentials (above synaptic reversal potential; Fig. 2c) to allow opening of NMDA receptors, inactivate voltage-sensitive Ca^{2+} channels (Fig. B in Supplementary Information), and minimize the effects of possible synaptic voltage escape. Under these conditions, fluorescence transients were unaffected by the selective

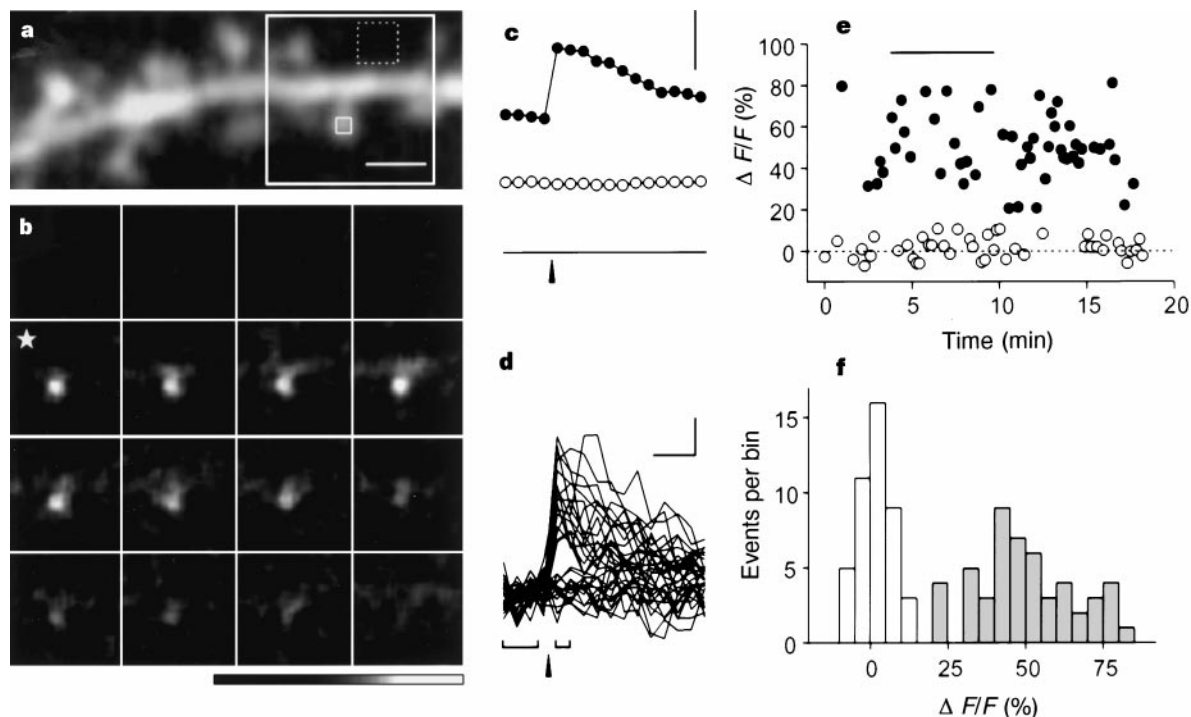


Figure 1 Probabilistic activation of NMDA receptors monitored by imaging spine $[Ca^{2+}]$ transients. **a**, Fluorescence image of tertiary apical dendritic branch of a CA1 pyramidal neuron. Scale bar: $2\ \mu\text{m}$. **b**, Sequence of frames (64 ms per frame, area shown by large box in **a**) showing relative change in fluorescence ($\Delta F/F(t)$) during synaptic stimulation (asterisk shows stimulus time). Grey scale: -5 to 90% $\Delta F/F$. **c**, Average fluorescence ($F(t)$) measured at spine (small solid box, filled

symbols) and background (dashed box, open symbols). Arrowhead indicates stimulation time; thin line indicates 0 level. Scale bar: 20 arbitrary units. **d**, 34 spine responses collected in the period indicated by horizontal bar in **e**. Scale bars: 25% $\Delta F/F$, 200 ms. **e**, Amplitude of fluorescence transients showing failures (open symbols) and releases (filled symbols). **f**, Corresponding histogram of response amplitudes.

AMPA (β -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor antagonist NBQX (2,3-dihydroxy-6-nitro-7-sulphamoylbenzo quinoxaline) (5 – $10\ \mu\text{M}$) and blocked by the NMDA-R antagonist D-CPP ($R(-)$ -3-(2-carboxypiperazine-4-yl)-propyl-1-phosphonic acid) ($10\ \mu\text{M}$; $4.2 \pm 3.5\%$ of control, $n = 5$). Fluorescence transients peaked at around 50 – 100 ms after a stimulus (Fig. 2b), consistent with the time course of the NMDA-R current⁵ (Fig. 2c), and decayed slowly (probably mainly reflecting unbinding of Ca^{2+} from the indicator¹¹). To test for a possible contribution of Ca^{2+} release from intracellular stores, we inhibited store refilling by using cyclopiazonic acid (CPA). In control experiments (see Methods), CPA treatment ($30\ \mu\text{M}$) increased the decay time of action-potential-evoked dendritic fluorescence transients to $167 \pm 11\%$ of control ($n = 3$, $P < 0.002$), showing that the drug was effective¹². CPA had no effect on the amplitude of synaptic fluorescence transients ($93 \pm 8\%$ of control ($n = 3$); compared to $95 \pm 4\%$ ($n = 5$) over a similar time period in untreated slices), showing that release from stores does not contribute to spine $[Ca^{2+}]$ transients under these conditions.

Synaptically activated spine fluorescence transients mediated by NMDA-Rs were large enough to distinguish successful transmitter release from failure of release (Fig. 1d–f), and could be monitored for up to ~ 500 trials (stimulation rate, $0.2\ \text{Hz}$). The probability of release measured at different synapses was broadly distributed (mean, 0.38 , s.d. = 0.23 ; $n = 32$). Fluorescence transients showed greater trial-to-trial amplitude variability than the background noise (s.d._{noise} = $5.1 \pm 0.4\%$ $\Delta F/F$, s.d._{resp} = $15.9 \pm 1\%$ $\Delta F/F$, noise-subtracted coefficient of variation (c.v.) = 0.34 ± 0.02 , $n = 27$; $\Delta F/F$ is the relative fluorescence). Failure and response distributions were well separated (on average, $2 \times \text{s.d.}_{\text{noise}} < \text{mean}_{\text{resp}} - 2 \times \text{s.d.}_{\text{resp}}$). Consistent with previous

electrophysiological measurements^{13,14,16,17}, responses showed pronounced paired-pulse facilitation (PPF) that was inversely related to release probability. The probability of at least one release increased to 0.75 ± 0.22 for a pair of stimuli $10\ \text{ms}$ apart, implying PPF = 2.08 ± 0.87 (see Methods).

To test for receptor saturation, we compared NMDA-R-spine $[Ca^{2+}]$ transients produced by single stimuli to those produced by pairs of stimuli separated by $10\ \text{ms}$ (Figs 2, 3). At this delay, glutamate that bound to NMDA-Rs during the first release will still occupy the receptors, owing to its slow unbinding rate^{5,6} (Fig. 2c). If receptors are saturated by a single release of glutamate then there will be no unbound receptors available for opening by a subsequent release. The expected response for a stimulus pair would therefore be about the same as for one stimulus. In fact, the average fluorescence transients evoked by pairs of stimuli were significantly larger than those evoked by single stimuli (Figs 2b, 3a, b) and the distributions of responses to stimulus pairs were shifted to larger amplitudes (Fig. 3c). Quantal peaks were not apparent in the distributions of paired stimulus responses, probably because of the large variance of single quantal NMDA-R responses. Similar shifts in response amplitudes were observed in all synapses examined (Fig. 3d; single stimulus response $r_1 = 43.8 \pm 2.8\%$ $\Delta F/F$, paired stimulus response $r_{1+2} = 55.1 \pm 3.2\%$ $\Delta F/F$, $n = 27$, $P < 10^{-9}$, paired sign test). Glutamate released following a single action potential therefore does not saturate postsynaptic NMDA-Rs.

What fraction of NMDA-Rs remain available for binding after one release event? Owing to the slow rise time of $[Ca^{2+}]$ transients mediated by NMDA-Rs (Fig. 2b), we could not distinguish between instances of single and double transmitter release in response to pairs of stimuli. Instead, we used a simple model to estimate the

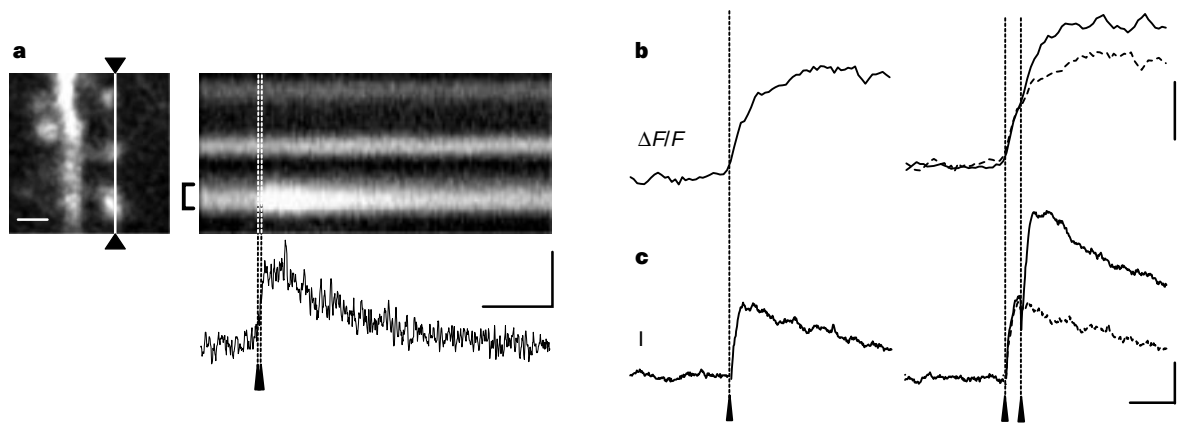


Figure 2 Time course of NMDA-R currents and spine fluorescence transients. Single or pairs of stimuli were delivered on alternating trials. **a** (left), Vertical line indicates the placement of the line scan (scale bar: 1 μm). **a** (right), Line scan in which vertical dimension corresponds to the image (left) and the horizontal dimension represents time. Stimulus times indicated by arrowheads. Below: $\Delta F/F(t)$ at the bottom spine (averaged over window indicated by brackets). Scale

bars: 250 ms, 50% $\Delta F/F$. **b**, **c**, Average EPSC (vertical bar, left in **c**) and corresponding $\Delta F/F(t)$ (**b**) for single (left) and double (right) stimuli at the same spine as **a**. The time base in **a** is expanded in **b** and **c**. The EPSC (and hence peak open probability) peaks at ~ 10 ms (when the second stimulus is delivered). Scale bars: 25% $\Delta F/F$, 100 pA, 25 ms.

number of double releases based on the failure rates for single and paired stimuli (see Methods). We could then obtain a prediction for the ratio of paired to single response amplitudes (r_{1+2}/r_1 ; both excluding failures), assuming linear summation. We assume a single site releasing either zero or one vesicle per action potential^{15,16} and that release probability on the second stimulus is independent of whether release occurred on the first. For most synapses, r_{1+2}/r_1 fell close to the value predicted for linear summation and well above the line of saturation ($r_{1+2}/r_1 = 1$) (Fig. 4a). Thus, NMDA-R-mediated $[\text{Ca}^{2+}]$ responses sum close to linearly, with the second response of a pair producing 0.8 ± 0.08 times the first response (Fig. 4b), translating into a fractional occupancy of less than 0.56 for a single vesicle.

There are several uncertainties in this calculation. First, we may underestimate the amplitude of larger responses because Ca^{2+} indicators have sublinear sensitivity to larger Ca^{2+} changes. Second, we may overestimate the expected number of double releases because the probability of release may be temporarily depressed following a successful release^{16,17}. Both these factors bias the calculation towards saturation. On the other hand, glutamate from more than one vesicle (released either by one presynaptic bouton or by several neighbouring boutons) may interact with the same population of NMDA-Rs (multi-vesicular release¹⁸). In this case, the quantal content due to the second action potential would be greater than assumed, biasing the calculation against saturation. Finally, although CA1 spines receive a single presynaptic contact^{8,9}, it is possible that some synapses could activate two independent populations of receptors. We consider this unlikely, as it would require perforated synapses with multivesicular release and very effective mechanisms to prevent glutamate diffusion between adjacent receptor populations.

We have shown that synaptic $[\text{Ca}^{2+}]$ transients in individual dendritic spines can be used to detect quantal release of neurotransmitter and quantify postsynaptic receptor opening. In contrast to purely electrical recordings, the number and identity of active synapses can be monitored functionally on a trial-by-trial basis. The signal-to-noise ratio of optically detected Ca^{2+} influx mediated by NMDA-Rs exceeds that of electrically detected NMDA-R currents and approaches that of electrically recorded AMPA-R currents. Control experiments demonstrated isolation of NMDA-R-mediated Ca^{2+} influx in our studies. Because the conditions of the intracellular environment were deliberately non-physiological (for example, high dye concentration, prolonged whole-cell washout

and depolarization), we do not rule out a physiological role for other Ca^{2+} sources.

NMDA-R-mediated spine $[\text{Ca}^{2+}]$ transients are not fully activated by the glutamate released by a single action potential. Rather, a second action potential occurring a short time (10 ms) later causes an additional increase in $[\text{Ca}^{2+}]$ at the same spine, implying an underlying increase in NMDA-R opening. The simplest interpretation of our findings is that synaptic NMDA-Rs are not saturated by

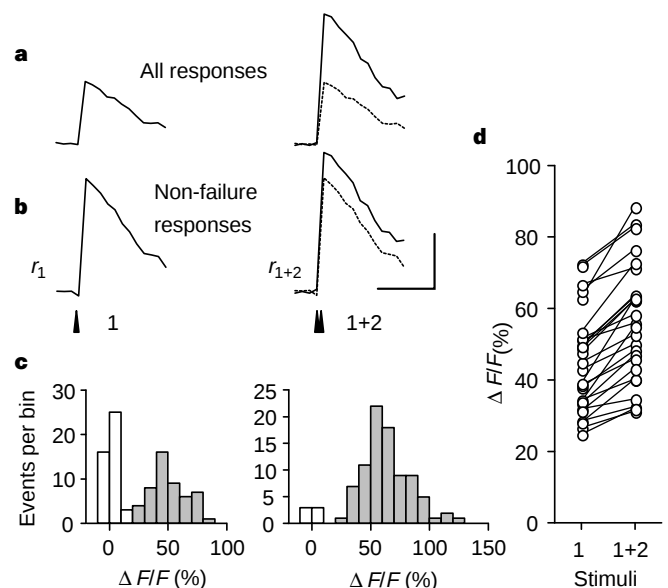


Figure 3 NMDA receptors are not saturated following a single action potential. **a–c**, Responses ($\Delta F/F(t)$) for one (left) and two (right) stimuli (10 ms interstimulus interval) at a single spine. **a**, Average of all responses including failures (failure rates: 0.46 and 0.07, respectively). **b**, Average of all responses excluding failures. At this spine, responses to double stimuli were significantly greater than to single stimuli ($r_1 = 50.3 \pm 2.29$ vs $r_{1+2} = 64.0 \pm 2.17\%$ $\Delta F/F$, $n = 50, 86$). **c**, Histograms showing the distributions of failure (white bars) and non-failure (shaded bars) response amplitudes. Note shift in distribution to larger amplitudes with double stimuli. Scale bars: 500 ms, 25% $\Delta F/F$. **d**, Average response amplitudes, excluding failures, for all spines tested ($n = 27$). For each spine, the corresponding average single and double stimulus response amplitudes are connected.

glutamate from a single vesicle. This finding is surprising because NMDA-Rs have a high affinity for glutamate and are thought to be outnumbered by a factor of 10–100 by transmitter molecules contained in one vesicle². Studies using kinetic analysis of competitive NMDA-R antagonists⁶ and biophysical modelling^{19,20} indicated that NMDA-R occupancy could approach 100% after a single action potential (but see ref. 21). Our data indicate that the concentration profile of glutamate reaching NMDA-Rs could be lower or briefer than conventional estimates. The higher temperature in our experiments may allow more effective glutamate clearance, possibly explaining differences between our study and previous ones^{6,18}. It could also be that some receptors directly opposite the release site are saturated while other receptors located on the same spine but further from the release site are not.

The absence of saturation of NMDA-Rs has a number of implications for our understanding of the function of excitatory synapses. Lack of saturation implies that the size of NMDA-R responses is limited by the transient concentration of glutamate, rather than the number of functional receptors. Changes in vesicular release or uptake mechanisms may therefore contribute to determining the amplitude and variability of postsynaptic responses. With uptake sufficiently efficient to prevent saturation, crosstalk or spillover between neighbouring synapses²³ would be less prominent than postulated in models assuming saturation²⁰. Indeed, even the closest neighbouring spines monitored in our study ($\sim 1 \mu\text{m}$ apart) acted independently, with no evidence of crosstalk (Fig. A of Supplementary Information). If AMPA-Rs are co-localized with NMDA-Rs, they are also likely not to be saturated under similar conditions^{18,24,25}.

Activation of NMDA-Rs and consequent Ca^{2+} entry into postsynaptic spines are critical to the plasticity of synaptic transmission at many central neurons²⁶. The degree of NMDA-R activation can determine whether a synapse undergoes long-term depression or potentiation²⁷, and the amount of NMDA-R opening depends on membrane potential²⁸ and hence postsynaptic activity. We have shown that NMDA-R opening is also sensitive to the number of presynaptic quanta released in a short time period. Thus, bursts of presynaptic action potentials may favour induction of potentiation not only by enhancing transmitter release and generally increasing postsynaptic depolarization, but also by enhancing fractional

NMDA-R opening and thereby increasing $[\text{Ca}^{2+}]$ transients at specific spines. Without saturation, changes in the concentration or rate of clearance of glutamate from the synaptic cleft might also alter the fractional activation of NMDA-Rs by single action potentials and hence modulate plasticity. $\square$

Methods

Preparation and electrophysiology. Rat (postnatal day 16–19) hippocampal slices (300–400 μm) were prepared in accordance with the animal care and use guidelines of CSHL. Slices were prepared using a chilled (2–5 °C) cutting solution containing (in mM) 110 choline-chloride, 25 NaHCO_3 , 25 D-glucose, 11.6 sodium ascorbate, 7 MgSO_4 , 3.1 sodium pyruvate, 2.5 KCl, 1.25 NaH_2PO_4 and 0.5 CaCl_2 , and then incubated in physiological saline (127 NaCl, 25 NaHCO_3 , 25 D-glucose, 2.5 KCl, 2 CaCl_2 , 1 MgSO_4 , 1.25 NaH_2PO_4 , 0.1 picrotoxin) for 30–60 min at 35 °C and then at room temperature until required. NBQX (5–10 μM) was present in about half of the experiments (otherwise a cut was made between CA3 and CA1). Experiments were done at 35 °C. Whole-cell patch electrodes (2–4 M Ω) contained 135 caesium methylsulphonate, 10 HEPES, 10 sodium phosphocreatine, 4 MgCl_2 , 4 Na ATP, 0.4 Na-GTP, and 0.2–0.4 Oregon Green BAPTA-1 or 2 (OGB-1, 2). Transmission was evoked by 0.1-ms voltage pulses delivered extracellularly by a glass pipette (2–3 μm tip)¹⁰. The stimulus intensity was adjusted to minimize failure rate at a target synapse. Control experiments with NBQX, D-CPP and CPA used the same alternating single/double pulse protocol as that used for testing NMDA-R saturation.

Two-photon imaging. We used a custom 2PLSM microscope¹⁰ (software: Lucent Technologies) to image fluorescence transients in CA1 spines. During imaging, the focus was periodically adjusted to correct for drift. Only synapses in which stable responses (determined by plots of $\Delta F/F$) from well-isolated Ca^{2+} sources could be verified (Fig. A of Supplementary Information) were analysed. Relative fluorescence changes were computed as $\Delta F/F(t) = (F(t) - F_0)/F_0$, where F_t is the fluorescence at time t . Baseline fluorescence (F_0) was determined by averaging the four frames (or 240 ms in the line scan) preceding the stimulus. Because background fluorescence was dominated by exogenous dye, it was not subtracted. We averaged the first two frames (50–100 ms in line scan) following the stimulus to calculate the amplitude. The relatively slow timescale over which spine $[\text{Ca}^{2+}]$ gradients dissipate²⁹ facilitated isolation of signals from neighbouring spines (Fig. A of Supplementary Information). The number of trials used to estimate response amplitudes and release probabilities was (median) 77 in frame mode ($n = 26$) and 26 in line-scan mode ($n = 4$). We assume that the influx of Ca^{2+} ions into a spine is proportional to the fraction of doubly bound receptors. To maximize signal-to-noise and avoid possible nonlinearities due to saturation of endogenous Ca^{2+} buffers, we used indicator concentrations large enough to dominate Ca^{2+} binding³⁰. To test CPA efficacy on action potential $[\text{Ca}^{2+}]$ transients, a low dye concentration (40 μM OGB-1) was used so that the decay was dominated by extrusion and uptake rather than indicator unbinding.

Analysis. We classified transmission failures and successes by eye using the fluorescence response waveform and the corresponding image. The average response evoked by a stimulus pair is $\Sigma p_{ij} r_{ij}$, where i, j indicate failure (=0) or success (=1) of the first (i) or second (j) stimulus, and p_{ij} and r_{ij} are the probabilities and average amplitudes for a given outcome. To calculate response linearity and occupancy, we assume a single release site^{8,9} releasing 0 or 1 vesicles per action potential^{15,16}; thus $r_1 = r_{01} = r_{10}$, where r_1 is the average non-failure response amplitude for a single stimulus. We define a metric of response linearity, $\Delta r_{11} = (r_{11} - r_1)/r_1$, the fractional amplitude of a response conditioned by a preceding release. $\Delta r_{11} = 0$ implies highly sublinear summation and therefore a high fractional occupancy of receptors after the first release. $\Delta r_{11} \approx 1$ implies linear summation and therefore low fractional receptor occupancy after the first release. Δr_{11} can be calculated from the non-failure responses to single stimuli (r_1) and to pairs of stimuli (r_{1+2}); according to the definition above, $r_{1+2} = (p_{10}r_1 + p_{01}r_1 + p_{11}r_{11})/(1 - p_{00})$. We further assume that the result of the first stimulus has no effect on the probability of release of the second stimulus. The apparent paired-pulse facilitation is then $\text{PPF} = (1 - p_{00}/p_0)/(1 - p_0)$, where p_0 and p_{00} are the measured failure rates for single and double stimuli, respectively. The ratio r_{1+2}/r_1 expected for the saturating (sat, where $r_{11} = r_1$) and linear (lin, where $r_{11} = 2r_1$) cases, in

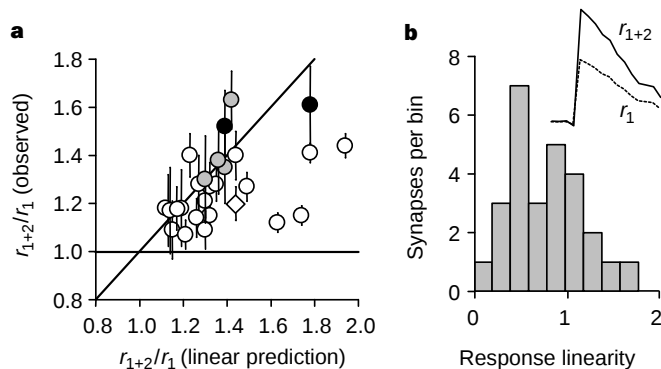


Figure 4 Quantification of NMDA-R response linearity (see Methods for calculation). **a**, Ratios of single to double stimulus non-failure responses, r_{1+2}/r_1 , plotted against the predicted ratios assuming linear response summation for each synapse tested. Most synapses are well above saturation (horizontal dashed line). Symbols designate indicator and imaging mode: OGB-1 frame mode (open circles); OGB-2 frame mode (filled circles); OGB-1 line scan mode (grey circles). The diamond corresponds to the synapse represented in Fig. 3a. **b**, NMDA-R response linearity, Δr_{11} . Linearity (absence of saturation) corresponds to a value of $\Delta r_{11} = 1$; saturation corresponds to $\Delta r_{11} = 0$. Inset, comparison of the average relative amplitudes of single (r_1) and calculated double releases ($r_{1+2} = 1.80r_1$).

terms of p_0 and p_{00} , is: $(r_{1+2}/r_1)^{\text{lin}} = (2 - p_0 - p_{00}/p_0)/(1 - p_{00})$ and $(r_{1+2}/r_1)^{\text{sat}} = 1$. Using the same model, we calculate $\Delta r_{11} = [r_{1+2}(1 - p_{00}) - r_1(1 - p_0)]/(1 - p_0 - p_{00}/p_0 + p_{00})$. As receptor occupancy following the second stimulus is ≤ 1 , we can estimate receptor occupancy following the first stimulus to be $\leq 1/(1 + \Delta r_{11})$. The calculation of receptor occupancy does not depend on the origin of release failures and is therefore insensitive to possible unreliability of axonal stimulation. All measurements are given as mean $\pm$ s.e.m. unless otherwise noted. The error bars for the observed ratio r_{1+2}/r_1 (Fig. 4a) were calculated by s.e.m. = $\sqrt{s_1^2 + s_2^2}$, $s_i = \mu_i/(\sigma_i\sqrt{n_i - 1})$, where μ_i , σ_i and n_i are, respectively, the mean, s.d. and number of observations of single and double responses.

Received 23 December 1998; accepted 3 March 1999.

- Denk, W. & Svoboda, K. Photon ummanship: why multiphoton imaging is more than a gimmick. *Neuron* **18**, 351–357 (1997).
- Frerking, M. & Wilson, M. Saturation of postsynaptic receptors at central synapses? *Curr. Opin. Neurobiol.* **6**, 395–403 (1996).
- Tang, C. M., Margulis, M., Shi, Q. Y. & Fielding, A. Saturation of postsynaptic glutamate receptors after quantal release of transmitter. *Neuron* **13**, 1385–1393 (1994).
- Auger, C., Kondo, S. & Marty, A. Multivesicular release at single functional synaptic sites in cerebellar stellate and basket cells. *J. Neurosci.* **18**, 4532–4547 (1998).
- Lester, R. A., Clements, J. D., Westbrook, G. L. & Jahr, C. E. Channel kinetics determine the time course of NMDA receptor-mediated synaptic currents. *Nature* **346**, 565–567 (1990).
- Clements, J. D., Lester, R. A., Tong, G., Jahr, C. E. & Westbrook, G. L. The time course of glutamate in the synaptic cleft. *Science* **258**, 1498–1501 (1992).
- Yuste, R. & Denk, W. Dendritic spines as basic functional units of neuronal integration. *Nature* **375**, 682–684 (1995).
- Harris, K. M. & Stevens, J. K. Dendritic spines of CA1 pyramidal cells in the rat hippocampus: Serial electron microscopy with reference to their biophysical characteristics. *J. Neurosci.* **9**, 2982–2997 (1989).
- Schikorski, T. & Stevens, C. F. Quantitative ultrastructural analysis of hippocampal excitatory synapses. *J. Neurosci.* **17**, 5858–5867 (1997).
- Mainen, Z. F. et al. Two-photon imaging in living brain slices. *Methods: A Companion to Methods in Enzymology* **18** (in the press).
- Regehr, W. G. & Atluri, P. P. Calcium transients in cerebellar granule cell presynaptic terminals. *Biophys. J.* **68**, 2156–2170 (1995).
- Markram, H., Helm, P. J. & Sakmann, B. Dendritic calcium transients evoked by single back-propagating action potentials in rat neocortical pyramidal neurons. *J. Physiol. (Lond.)* **485**, 1–20 (1995).
- Manabe, T. & Nicoll, R. A. Long-term potentiation: evidence against an increase in transmitter release probability in the CA1 region of the hippocampus. *Science* **265**, 1888–1892 (1994).
- Mainen, Z. F., Jia, Z. P., Roder, J. & Malinow, R. Use-dependent AMPA receptor block in mice lacking GluR2 suggests postsynaptic site for LTP expression. *Nature Neurosci.* **1**, 579–586 (1998).
- Korn, H., Mallet, A., Triller, A. & Faber, D. S. Transmission at a central inhibitory synapse. II. Quantal description of release, with a physical correlate for binomial n. *J. Neurophysiol.* **48**, 679–707 (1982).
- Stevens, C. F. & Wang, Y. Facilitation and depression at single central synapses. *Neuron* **14**, 795–802 (1995).
- Dobrunz, L. E., Huang, E. P. & Stevens, C. F. Very short-term plasticity in hippocampal synapses. *Proc. Natl Acad. Sci. USA* **94**, 14843–14847 (1997).
- Tong, G. & Jahr, C. E. Multivesicular release from excitatory synapses of cultured hippocampal neurons. *Neuron* **12**, 51–59 (1994).
- Clements, J. D. Transmitter timecourse in the synaptic cleft: its role in central synaptic function. *Trends Neurosci.* **19**, 163–171 (1996).
- Rusakov, D. A. & Kullmann, D. M. Extrasynaptic glutamate diffusion in the hippocampus: ultrastructural constraints, uptake, and receptor activation. *J. Neurosci.* **18**, 3158–3170 (1998).
- Holmes, W. R. Modeling the effect of glutamate diffusion and uptake on NMDA and non-NMDA receptor saturation. *Biophys. J.* **69**, 1734–1747 (1995).
- Tong, G. & Jahr, C. E. Block of glutamate transporters potentiates postsynaptic excitation. *Neuron* **13**, 1195–1203 (1994).
- Barbour, B. & Hausser, M. Intersynaptic diffusion of neurotransmitter. *Trends Neurosci.* **20**, 377–384 (1997).
- Faber, D. S., Young, W. S., Legendre, P. & Korn, H. Intrinsic quantal variability due to stochastic properties of receptor–transmitter interactions. *Science* **258**, 1494–1498 (1992).
- Silver, R. A., Cull-Candy, S. G. & Takahashi, T. Non-NMDA glutamate receptor occupancy and open probability at a rat cerebellar synapse with single and multiple release sites. *J. Physiol. (Lond.)* **494**, 231–250 (1996).
- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Cummings, J. A., Mulkey, R. M., Nicoll, R. A. & Malenka, R. C. Ca^{2+} signaling requirements for long-term depression in the hippocampus. *Neuron* **16**, 825–833 (1996).
- Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochiantz, A. Magnesium gates glutamate-activated channels in mouse central neurons. *Nature* **307**, 462–465 (1984).
- Svoboda, K., Tank, D. W. & Denk, W. Direct measurement of coupling between dendritic spines and shafts. *Science* **272**, 716–719 (1996).
- Helmchen, F., Imoto, K. & Sakmann, B. Ca^{2+} buffering and action potential-evoked Ca^{2+} signaling in dendrites of pyramidal neurons. *Biophys. J.* **70**, 1069–1081 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. This work was supported by a Burroughs Wellcome Career Award (Z.F.M.), grants from NIH and the Mather's Foundation (R.M.), and grants from the Pew and Whitaker Foundations (K.S.).

Correspondence and requests for materials should be addressed to K.S. (e-mail: svoboda@cshl.org). A movie of spine Ca^{2+} transients is available on the authors' website (<http://www.cshl.org/public/SCIENCE/spine.gif>).

Ca^{2+} /calmodulin binds to and modulates P/Q-type calcium channels

Amy Lee, Scott T. Wong, Daniel Gallagher, Bin Li, Daniel R. Storm, Todd Scheuer & William A. Catterall

Department of Pharmacology, University of Washington, Seattle, Washington 98195-7280, USA

Neurotransmitter release at many central synapses is initiated by an influx of calcium ions through P/Q-type calcium channels^{1,2}, which are densely localized in nerve terminals³. Because neurotransmitter release is proportional to the fourth power of calcium concentration^{4,5}, regulation of its entry can profoundly influence neurotransmission. N- and P/Q-type calcium channels are inhibited by G proteins^{6,7}, and recent evidence indicates feedback regulation of P/Q-type channels by calcium⁸. Although calcium-dependent inactivation of L-type channels is well documented^{9–11},

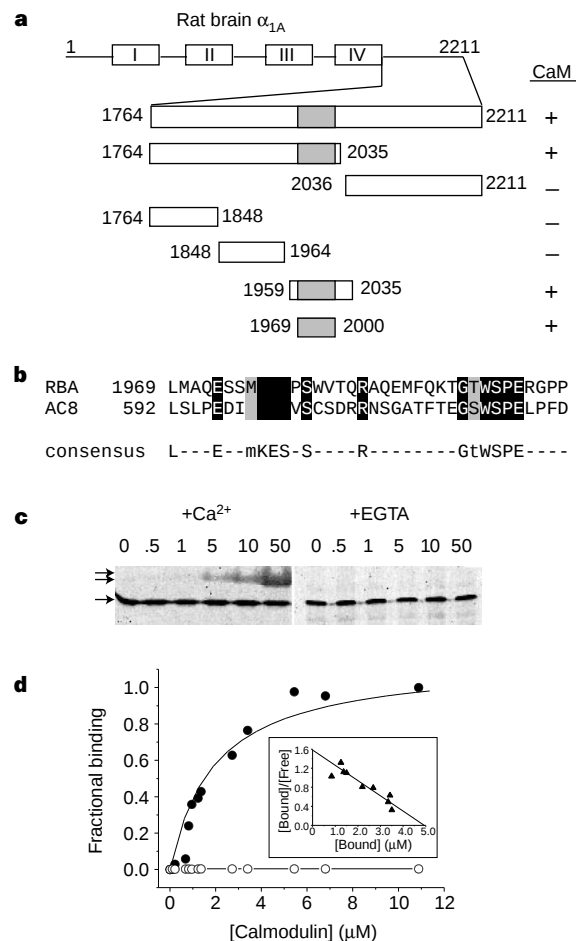


Figure 1 Calmodulin binding to the C-terminal domain of the α_{1A} subunit. **a**, Identification of a calmodulin-binding domain (α_{1A} CBD, shaded region) in α_{1A} . Constructs delimited by the indicated amino acids that interacted with calmodulin (CaM) in the yeast two-hybrid assay are indicated (+). **b**, Alignment of the α_{1A} CBD (RBA) with type 8 adenylyl cyclase (AC8). **c**, Ca^{2+} -dependent binding of α_{1A} CBD peptide (amino acids 1969–2000) to calmodulin in gel-shift assays with the indicated peptide:calmodulin molar ratios. Single arrow, unbound calmodulin; double arrows, peptide-bound calmodulin. **d**, Affinity of α_{1A} CBD for calmodulin estimated by fluorescence anisotropy ($K_d = 1.95 \pm 0.53 \mu\text{M}$). Closed circles, α_{1A} CBD peptide; open circles, control peptide. Inset, Scatchard plot.

little is known about how calcium modulates P/Q-type channels. Here we report a calcium-dependent interaction between calmodulin and a novel site in the carboxy-terminal domain of the α_{1A} subunit of P/Q-type channels. In the presence of low concentrations of intracellular calcium chelators, calcium influx through P/Q-type channels enhances channel inactivation, increases recovery from inactivation and produces a long-lasting facilitation of the calcium current. These effects are prevented by over-expression of a calmodulin-binding inhibitor peptide and by deletion of the calmodulin-binding domain. Our results reveal an unexpected association of Ca^{2+} /calmodulin with P/Q-type calcium channels that may contribute to calcium-dependent synaptic plasticity.

To identify proteins that influence P/Q-type calcium-channel function through direct protein-protein interactions, we performed a yeast two-hybrid screen of a rat brain complementary DNA library using the intracellular C-terminus of the rat α_{1A} subunit (rbA; ref.12) as bait. We isolated multiple interacting clones that contained a 1.05 kilobase (kb) cDNA encoding rat calmodulin. These calmodulin cDNAs sustained cell growth in the yeast two-hybrid assay in combination with the α_{1A} cDNA, but not when cDNA encoding the corresponding region of the rat α_{1B} subunit of N-type channels was used (not shown). Deletion analysis revealed a 32-amino-acid calmodulin-binding domain (α_{1A} CBD) that was sufficient for the interaction with calmodulin (Fig. 1a), whereas neighbouring sequences could not bind calmodulin. α_{1A} CBD lacks structural similarity to previously identified calmodulin-binding domains, but it has a strong similarity to a region in the Ca^{2+} /calmodulin-stimulated adenylyl cyclase type 8 (AC8, Fig. 1b). Peptides corresponding to α_{1A} CBD and the AC8 domain inhibit calmodulin stimulation of adenylyl cyclase activity (not shown). Thus, amino acids conserved between α_{1A} CBD and AC8 may represent unique structural determinants for calmodulin binding.

To characterize the interaction with calmodulin further, a synthetic peptide corresponding to α_{1A} CBD (Leu 1969 to Glu 2000) was analysed in a gel-shift assay in which the larger peptide-bound form of calmodulin should migrate with a lower mobility under non-denaturing conditions¹³. In the presence of calcium, a higher-molecular-mass band representing the peptide/calmodulin complex was observed with increasing concentrations of the α_{1A} CBD

peptide (Fig. 1c, double arrow), but not with a peptide derived from a different region of α_{1A} (CNA-1 (ref. 3); not shown) or when the calcium buffer EGTA was substituted for calcium (Fig. 1c). Binding of calmodulin to α_{1A} CBD was analysed further in fluorescence anisotropy studies, which measured the increase in polarization of the fluorescence of the two tryptophan residues in the α_{1A} CBD peptide upon binding to Ca^{2+} /calmodulin (Fig. 1d). Both α_{1A} CBD and a control peptide exhibited characteristic tryptophan fluorescence at 360 nm when excited at 295 nm. Incubation of calmodulin with α_{1A} CBD, but not the control peptide, caused a progressive increase in fluorescence polarization, indicating saturable binding of calmodulin with a dissociation constant of 2 μM in 1 mM Ca^{2+} (Fig. 1d).

Hexahistidine-tagged fusion proteins containing α_{1A} CBD immobilized on Ni^{2+} -agarose pulled down calmodulin in the presence, but not the absence, of calcium (Fig. 2a). Similarly, P/Q-type Ca^{2+} channels containing full-length α_{1A} subunits expressed in the human embryonic kidney cell line tsA-201 bound specifically to calmodulin immobilized on agarose beads in the absence, but not the presence, of excess free calmodulin (Fig. 2b). Calcium-dependent association of calmodulin with α_{1A} subunits co-expressed in tsA-201 cells was also observed in co-immunoprecipitation studies (Fig. 2c). Calmodulin co-immunoprecipitated with α_{1A} subunits only when intracellular Ca^{2+} was raised by the calcium ionophore A23187 (5 μM) (Fig. 2c, lane 4). Deletion of the α_{1A} CBD ($\alpha_{1A\Delta\text{CBD}}$, amino acids 1969–2060) prevented the co-immunoprecipitation (Fig. 2c, lane 6).

Ion-channel function is modulated by direct interaction with calmodulin^{14–16}. To examine the functional impact of calmodulin binding to α_{1A} , we analysed transfected P/Q-type channels in whole-cell patch-clamp recordings of tsA-201 cells using intracellular solutions containing low (0.5 mM) and high (10 mM) concentrations of EGTA to differentially buffer entering Ca^{2+} , or with Ba^{2+} as the permeant ion to prevent effects requiring binding to

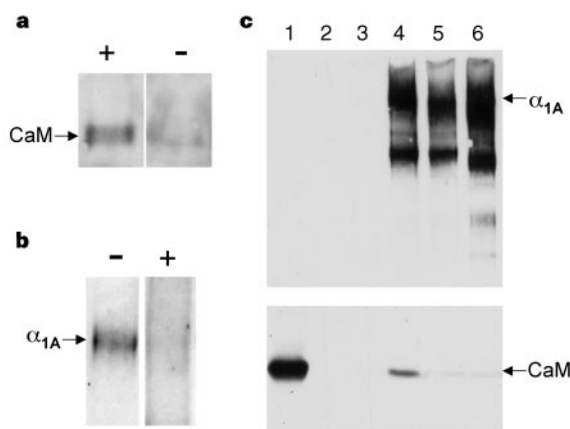


Figure 2 Ca^{2+} -dependent binding of calmodulin to α_{1A} . **a**, *In vitro* binding of calmodulin to His-tagged α_{1A} (1959–2035) with (+) or without (–) Ca^{2+} . **b**, Immunoblot of full-length α_{1A} expressed in tsA-201 cells and specifically bound to calmodulin/agarose with (+) or without (–) excess free calmodulin. **c**, Ca^{2+} -dependent association of calmodulin and full-length α_{1A} . Ca^{2+} channels were immunoprecipitated and analysed by immunoblotting for α_{1A} (upper panel) or calmodulin (lower panel). Lanes: 1, calmodulin (1 μg); 2, mock transfection; 3, control IgG; 4, complete assay system after stimulation with A23187; 5, complete assay system but with no A23187; 6, complete assay system but with $\alpha_{1A\Delta\text{CBD}}$ deletion mutant.

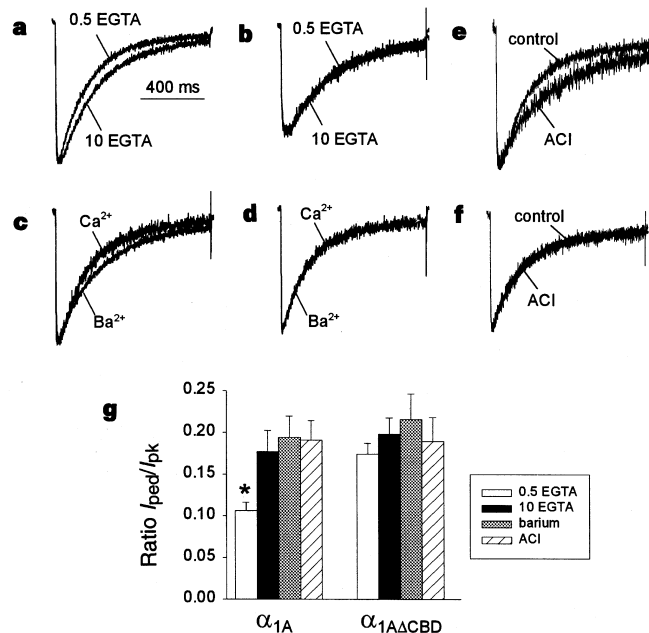


Figure 3 Effects of Ca^{2+} and calmodulin on inactivation of P/Q-type channels. **a–f**, Normalized Ca^{2+} currents elicited by a 1 s test pulse to +10 mV from a holding potential of –80 mV. Records are from cells expressing α_{1A} (**a**, **c**, **e**) or $\alpha_{1A\Delta\text{CBD}}$ (**b**, **d**, **f**). Cells in **e** and **f** were co-transfected with a calmodulin inhibitor peptide (ACI) or control vector (control). Intracellular recording solutions contained 0.5 mM EGTA (**a**, **b**, upper traces; **e**, **f**) or 10 mM EGTA (**a**, **b**, lower traces); extracellular solutions contained Ca^{2+} (**a**, **b**, **c**, **d**, upper traces; **e**, **f**) or Ba^{2+} (**c**, **d**, lower traces). **g**, Loss of Ca^{2+} -dependent effects on inactivation in $\alpha_{1A\Delta\text{CBD}}$ -expressing cells. $I_{\text{ped}}/I_{\text{pk}}$ values are represented as mean $\pm$ s.e.m. Asterisk, $P < 0.01$.

calmodulin¹⁷. With 0.5 mM intracellular EGTA, the calcium current I_{Ca} inactivated faster ($\tau = 186.6 \pm 12.0$ ms, $n = 15$) than in 10 mM EGTA ($\tau = 323.8 \pm 39.0$ ms, $n = 7$; $P < 0.01$) or when Ba^{2+} was used as the permeant ion ($\tau = 285 \pm 30$ ms, $n = 11$, $P < 0.01$; Fig. 3a, c). The extent of inactivation during the test pulse was quantified by comparing the current amplitude at the end of a test pulse (I_{ped}) to the peak current (I_{pk}). In 0.5 mM EGTA, I_{Ca} inactivated more completely ($I_{ped}/I_{pk} = 0.11 \pm 0.01$) than in 10 mM EGTA ($I_{ped}/I_{pk} = 0.18 \pm 0.02$; $P < 0.01$) or with Ba^{2+} in the extracellular solution ($I_{ped}/I_{pk} = 0.19 \pm 0.03$; $P < 0.01$) (Fig. 3g). To test whether Ca^{2+} /calmodulin was important for these effects, we co-expressed a high-affinity calmodulin-binding peptide from adenylyl cyclase type 1 (ref. 18) to compete with calcium channels for binding to calmodulin. I_{Ca} inactivated less in cells co-expressing the calmodulin inhibitor peptide ($I_{ped}/I_{pk} = 0.19 \pm 0.02$ versus 0.11 ± 0.01 ; $P < 0.05$, $n = 8$; Fig. 3e, g). Furthermore, inactivation of P/Q-type calcium channels containing $\alpha_{1A\Delta CBD}$ subunits was not significantly different with 0.5 mM EGTA ($I_{ped}/I_{pk} = 0.17 \pm 0.01$, $n = 14$), 10 mM EGTA ($I_{ped}/I_{pk} = 0.17 \pm 0.02$, $n = 7$), Ba^{2+} ($I_{ped}/I_{pk} = 0.22 \pm 0.03$, $n = 9$) or the calmodulin inhibitor peptide ($I_{ped}/I_{pk} = 0.19 \pm 0.03$, $n = 8$; Fig. 3b, d, f, g), even though voltage-dependent activation and inactivation of the mutant channel were similar to those of the wild type (not shown). Our data show that calcium entering through P/Q-type channels promotes inactivation through Ca^{2+} /calmodulin binding to the α_{1A} CBD, although other sequences may be involved in coupling calmodulin binding to channel regulation. This mechanism may contribute to the calcium dependence of the inactivation of P-type channels in the

presynaptic terminal of the calyx of Held⁸.

Because recovery from inactivation may be an important determinant of calcium influx and neurotransmitter release at synapses^{8,19,20}, we studied the effects of calcium and calmodulin on the recovery of I_{Ca} following a 3.5 s depolarizing stimulus. In 0.5 mM EGTA, I_{Ca} recovered with a single exponential time course ($\tau = 11.3 \pm 0.9$ s, $n = 12$) and showed a small but consistent facilitation compared to cells studied with 10 mM EGTA or Ba^{2+} (facilitation ratio = 1.15 ± 0.02 at 45 s after the prepulse, $P < 0.05$; Fig. 4a). This facilitation required calcium entry during the prepulse, as it was not observed following a prepulse to +100 mV, which would fully activate calcium channels but prevent calcium entry by reversing the electrochemical driving force (not shown). Recovery followed a biexponential time course and showed no facilitation for I_{Ca} in 10 mM EGTA and for I_{Ba} (Fig. 4a, c), and cells transfected with $\alpha_{1A\Delta CBD}$ showed no calcium-dependent enhancement of recovery from inactivation and no facilitation (Fig. 4b, d). Co-expression of the calmodulin inhibitor peptide caused a substantial slowing of I_{Ca} recovery for wild-type channels ($\tau = 14.5 \pm 1.6$ s, $n = 5$ for peptide treated versus $\tau = 9.6 \pm 1.5$ s, $n = 4$ for control; $P < 0.05$), reduced the extent of recovery and blocked the facilitation of the calcium current observed after long recovery times (Fig. 4e). In contrast, the inhibitor peptide had no effect on the recovery of I_{Ca} in cells expressing the mutant $\alpha_{1A\Delta CBD}$ ($\tau = 10.2 \pm 1.7$ s, $n = 5$ versus $\tau = 9.6 \pm 1.2$ s, $n = 5$; $P = 0.69$) (Fig. 4f).

Our results support a model in which calcium entering through P/Q-type channels promotes Ca^{2+} /calmodulin binding to the α_{1A} subunit. The association of Ca^{2+} /calmodulin with the channel

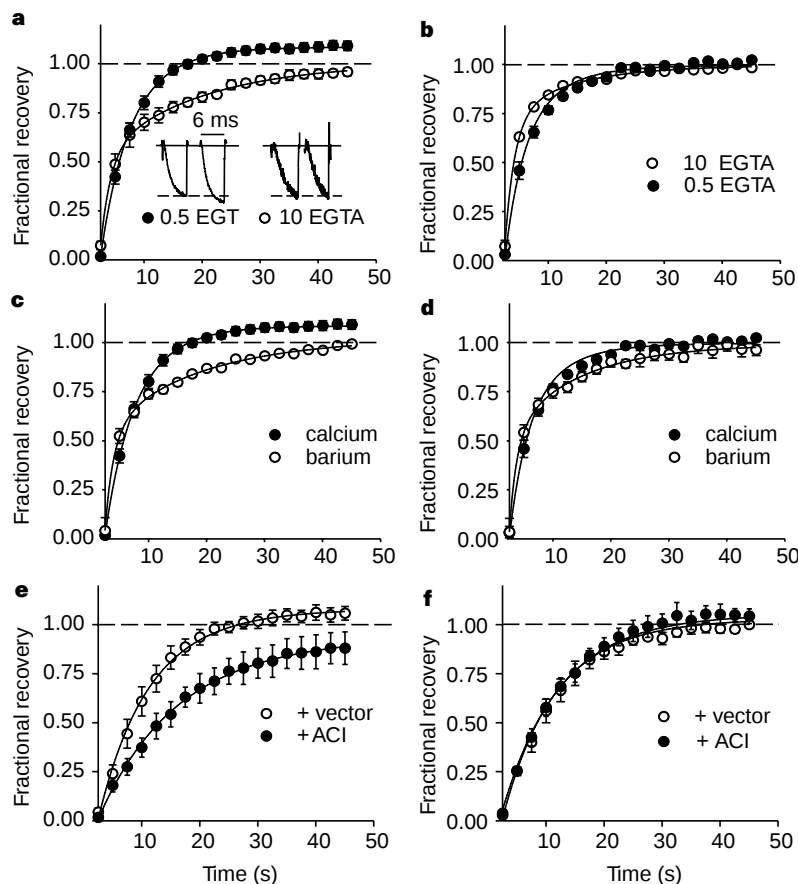


Figure 4 Effects of Ca^{2+} and calmodulin on recovery from inactivation. **a–f**, Fractional recovery was measured as the ratio of currents elicited by test pulses (6 ms, to +10 mV) applied before and every 2.5 s after an inactivating prepulse (3.5 s, to +10 mV). Dashed line indicates full recovery. Inset, representative normalized current traces before (left) and 45 s after (right) the prepulse.

Recordings were from cells expressing α_{1A} (**a**, **c**, **e**) or $\alpha_{1A\Delta CBD}$ (**b**, **d**, **f**). Intracellular recording solutions contained 0.5 mM EGTA and the extracellular divalent cation was Ca^{2+} except as indicated (**a–d**). Effect of overexpression of a calmodulin binding inhibitor peptide (ACI) or control vector (**e**, **f**).

accelerates inactivation, enhances recovery from inactivation and augments further calcium influx by facilitating the calcium current so that it is greater after full recovery from inactivation. These relatively slow modulatory effects, on the timescale of seconds, may interact to sharpen the response of P/Q-type channels to repeated trains of stimuli. Accumulation of intracellular calcium during a train of impulses would increase inactivation, shorten calcium transients and contribute to terminating the calcium signal following the train. However, binding of Ca^{2+} /calmodulin would also increase recovery from inactivation, thereby preparing the channels to respond more fully to a subsequent train of stimuli, and would enhance that response by facilitation of the calcium current. The overall effect would be briefer, more intense calcium signals in dendrites and nerve terminals with P/Q-type channels. As facilitation of I_{Ca} lasts for 15–45 s after the inactivating depolarization, these modulatory interactions between Ca^{2+} /calmodulin and P/Q-type channels in nerve terminals may contribute to calcium-dependent facilitation of synaptic transmission in response to repeated trains of depolarizing stimuli^{21,22,23}, including calcium-dependent re-filling of the readily releasable pool of synaptic vesicles²⁰. In contrast to the subtle effects on inactivation observed here, association of calmodulin with small conductance calcium-activated potassium channels is essential for their gating by calcium¹⁵, and calmodulin binding is required for calcium-dependent inactivation of L-type channels^{24,25}. The novel calmodulin binding site revealed here, which lacks an IQ domain²⁶, is now implicated in regulation of both calcium and cyclic AMP signalling pathways, based on its presence in P/Q-type channels and AC8. The potential physiological significance of this calmodulin-binding domain *in vivo* is illustrated by the severe ataxia, absence epilepsy and cerebellar degeneration observed in *leaner* mice, in which α_{1A} CBD is deleted by defective mRNA splicing²⁷.

Note added in proof: Following submission of this work for publication, three papers described Ca^{2+} -dependent inactivation of L-type Ca^{2+} channels caused by the interaction of calmodulin with an IQ motif in the C-terminal domain. They are: N. Qin, R. Olcese, M. Bransby, T. Lin and L. Birnbaumer, *Proc. Natl Acad. Sci.* **96**, 2435–2438 (1999); R. D. Zühlke, G. S. Pitt, K. Deisseroth, R. W. Tsien and H. Reuter, *Nature* **399**, 159–162 (1999); and B. Z. Peterson, C. D. DeMaria and D. T. Yue *Neuron* **22**, 549–558 (1999). □

Methods

Yeast two-hybrid analysis. cDNA encoding amino acids 1764–2211 of the rat brain α_{1A} subunit was generated by PCR and subcloned into pAS2-1 (Clontech). Reporter yeast strain Y190 bearing this plasmid was transformed with a rat brain cDNA library subcloned in pGAD10 (Clontech). Around 10^6 transformants were screened both for growth on medium lacking histidine and for β -galactosidase activity, as described²⁸. Deletion constructs were generated by PCR, subcloned into pAS2-1 and co-transformed with calmodulin cDNA in pGAD10 into Y190 to assess their interaction *in vivo*.

Gel-shift assays. Bovine brain calmodulin (300 pmol) was incubated with varying concentrations of a 32-amino-acid peptide corresponding to the α_{1A} CBD in 10 mM sodium-HEPES (pH 7.4) and 2 mM CaCl_2 or 2 mM EGTA for 15 min at room temperature. Bound complexes were resolved by non-denaturing PAGE and visualized by staining with Coomassie blue.

Fluorescence anisotropy. Fluorescence anisotropies r were derived from fluorescence polarization P measurements at 0° and 90° by the equations: $r = 2P/(3 - P)$ and $P = (I_{\parallel} - I_{\perp})/(I_{\parallel} + I_{\perp})$. Fractional binding was derived as $f_B = (r - r_F)/(0.21[r_B - r] + r - r_F)$ where the factor 0.21 accounts for the change in quantum yield upon peptide binding to calmodulin. All measurements were adjusted for the anisotropic contribution of calmodulin itself. A peptide corresponding to amino acids 1058–1081 of type 3 adenylyl cyclase was used as a negative control.

In vitro binding and co-immunoprecipitation. A hexahistidine (His)-tagged fusion protein containing amino acids 1959–2035 of α_{1A} was generated by PCR and subcloned in pTrcHisC (Invitrogen). The purified His fusion protein (3 μg) was incubated with calmodulin (10 μg) in the presence of 2 mM Ca^{2+} or

0.1 mM EGTA and subsequently with Ni^{2+} -agarose beads. Bound proteins were eluted with imidazole (600 mM) and analysed by immunoblotting with a monoclonal anti-calmodulin antibody (Upstate Biochemicals). For co-immunoprecipitation experiments, tsA-201 cells were transfected by the calcium-phosphate method with cDNAs encoding α_{1A} , β_{1b} and $\alpha_{2\delta}$ at 1:1:1 molar ratios. After at least 48 h, transfected cells were incubated for 15–30 min with or without A23187 (5 μM) in the presence of 2 mM CaCl_2 , solubilized in buffer A (20 mM Tris, pH 7.3, 150 mM NaCl, 1% NP40, 2 mM CaCl_2) and incubated with antibodies (anti-CNA5 specific for the Ca^{2+} -channel α_{1A} subunit²⁹). Immune complexes were isolated on protein A-Sepharose, resolved by SDS-PAGE, transferred to nitrocellulose membranes and analysed by immunoblotting.

For binding to immobilized calmodulin, a solubilized membrane fraction from cells transfected with Ca^{2+} -channel subunits as above was prepared by solubilization in buffer containing 1% Triton X-100 as described³⁰ and incubated with calmodulin bound to agarose beads in buffer A for 1 h at 4°C . Beads were washed in buffer A and bound proteins were eluted in buffer A containing 5 mM EGTA in place of CaCl_2 . Washes and elutions were concentrated and analyzed by SDS-PAGE and immunoblotting with anti-CNA5 antibodies.

Electrophysiological recordings. tsA-201 cells were co-transfected with P/Q-type Ca^{2+} -channel subunits as described above and a CD8 expression plasmid to allow visual identification of transfected cells with CD8-antibody-coated microspheres (Dynal). For some experiments, an expression construct containing the calmodulin-binding domain of type I adenylyl cyclase (amino acids 481–575, ref. 18) was generated by PCR and subcloned into pCEP4 (Invitrogen). This construct, or pCEP4 for control experiments, was cotransfected at a 5:1 molar ratio with Ca^{2+} -channel subunits. Ca^{2+} currents were recorded in whole-cell patch-clamp experiments as described⁶. Extracellular recording solutions contained 150 mM Tris, 1 mM MgCl_2 and 10 mM CaCl_2 or BaCl_2 . Intracellular solutions consisted of 120 mM *N*-methyl-D-glucamine, 60 mM HEPES, 1 mM MgCl_2 , 2 mM Mg-ATP and 0.5 or 10 mM EGTA. The pH of both solutions was adjusted to 7.3 with methanesulphonic acid.

Received 26 February; accepted 14 April 1999.

- Dunlap, K., Luebke, J. I. & Turner, T. J. Exocytotic Ca^{2+} channels in mammalian central neurons. *Trends Neurosci.* **18**, 89–98 (1995).
- Tsien, R. W., Elinor, P. T. & Horne, W. A. Molecular diversity of voltage-dependent calcium channels. *Trends Neurosci.* **12**, 349–354 (1991).
- Westenbroek et al. Immunohistochemical identification and subcellular distribution of the α_{1A} subunits of brain calcium channels. *J. Neurosci.* **15**, 6403–6418 (1995).
- Dodge, F. A. Jr & Rahamimoff, R. Co-operative action of calcium ions in transmitter release at the neuromuscular junction. *J. Physiol. (Lond.)* **193**, 419–432 (1967).
- Mintz, I. M., Sabatini, B. L. & Regehr, W. G. Calcium control of transmitter release at a cerebellar synapse. *Neuron* **15**, 675–688 (1995).
- Herlitze et al. Modulation of Ca^{2+} channels by G protein $\beta\gamma$ subunits. *Nature* **380**, 258–262 (1996).
- Ikeda, S. R. Voltage-dependent modulation of N-type calcium channels by G-protein $\beta\gamma$ subunits. *Nature* **380**, 255–258 (1996).
- Forsythe, I. D., Tsujimoto, T., Barnes-Davies, M., Cuttle, M. F. & Takahashi, T. Inactivation of presynaptic calcium current contributes to synaptic depression at a fast central synapse. *Neuron* **20**, 797–807 (1998).
- Neely, A., Olcese, R., Wei, X., Birnbaumer, L. & Stefani, E. Ca^{2+} -dependent inactivation of a cloned cardiac Ca^{2+} channel α_1 subunit (α_{1C}) expressed in *Xenopus* oocytes. *Biophys. J.* **66**, 1895–1903 (1994).
- Imredy, J. P. & Yue, D. T. Mechanism of Ca^{2+} -sensitive inactivation of L-type Ca^{2+} channels. *Neuron* **12**, 1301–1318 (1994).
- Höfer, G. F. et al. Intracellular Ca^{2+} inactivates L-type Ca^{2+} channels with a Hill coefficient of 1 and an inhibition constant of 4 μM by reducing channel's open probability. *Biophys. J.* **73**, 1857–1865 (1997).
- Starr, T. V. B., Prystay, W. & Snutch, T. P. Primary structure of a calcium channel that is highly expressed in the rat cerebellum. *Proc. Natl Acad. Sci. USA* **88**, 5621–5625 (1991).
- Erickson-Viitanen, S. & DeGrado, W. F. Recognition and characterization of calmodulin-binding sequences in peptides and proteins. *Methods Enzymol.* **139**, 455–478 (1987).
- Liu, M., Chen, T. Y., Ahamed, B., Li, J. & Yau, K. W. Calcium-calmodulin modulation of the olfactory cyclic nucleotide-gated cation channel. *Science* **266**, 1348–1354 (1994).
- Xia, X. M. et al. Mechanism of calcium gating in small-conductance calcium-activated potassium channels. *Nature* **395**, 503–507 (1998).
- Saimi, Y. & Kung, C. Ion channel regulation by calmodulin binding. *FEBS Lett.* **350**, 155–158 (1994).
- Teo, T. S. & Wang, J. H. Mechanism of activation of a cyclic adenosine 3':5'-monophosphate phosphodiesterase from bovine heart by calcium ions. *J. Biol. Chem.* **248**, 5950–5955 (1973).
- Wu, Z., Wong, S. T. & Storm, D. R. Modification of the calcium and calmodulin sensitivity of the type I adenylyl cyclase by mutagenesis of its calmodulin binding domain. *J. Biol. Chem.* **268**, 23766–23768 (1993).
- von Gersdorff, H. & Matthews, G. Calcium-dependent inactivation of calcium current in synaptic terminals of retinal bipolar neurons. *J. Neurosci.* **16**, 115–122 (1996).
- Wang, L. Y. & Kaczmarek, L. K. High-frequency firing helps replenish the readily releasable pool of synaptic vesicles. *Nature* **394**, 384–388 (1998).
- Kamiya, H. & Zucker, R. S. Residual Ca^{2+} and short-term synaptic plasticity. *Nature* **371**, 603–606 (1994).
- Cuttle, M. F., Tsujimoto, T., Forsythe, I. D. & Takahashi, T. Facilitation of the presynaptic calcium current at an auditory synapse in rat brainstem. *J. Physiol. (Lond.)* **512**, 723–729 (1998).
- Borst, J. G. G. & Sakmann, B. Facilitation of presynaptic calcium currents in the rat brainstem. *J. Physiol. (Lond.)* **513**, 149–155 (1998).
- Zühlke, R. D. & Reuter, H. Identification of a single amino acid residue as molecular determinant of calcium-dependent inactivation or facilitation of L-type calcium channels. *Biophys. J.* **76**, A343 (1999).
- Peterson, B. Z. et al. Individual amino acids within a consensus EF-hand motif are critical for calcium-dependent inactivation of α_{1C} calcium channels. *Biophys. J.* **76**, A340 (1999).

26. James, P., Vorherr, T. & Carafoli, E. Calmodulin-binding domains: just two-faced or multi-faceted? *Trends Biochem. Sci.* **20**, 38–42 (1995).
27. Fletcher, C. F. *et al.* Absence epilepsy in tottering mutant mice is associated with calcium channel defects. *Cell* **87**, 607–617 (1996).
28. Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805–816 (1993).
29. Sakurai, T., Westenbroek, R. E., Rettig, J., Hell, J. & Catterall, W. A. Biochemical properties and subcellular distribution of the BI and rBA isoforms of α_{1A} subunits of brain calcium channels. *J. Cell. Biol.* **134**, 511–528 (1996).
30. Yokoyama, C. T., Sheng, Z.-H. & Catterall, W. A. Phosphorylation of the synaptic protein interaction site on N-type calcium channels inhibits interactions with SNARE proteins. *J. Neurosci.* **17**, 6929–6938 (1997).

Acknowledgements. We thank T. Snutch, M. Harpold and K. Campbell for cDNAs encoding Ca^{2+} -channel subunits. This work was supported by an NRSA postdoctoral fellowship from the NIH to A.L. and by research grants from the NIH to D.R.S. and W.A.C.

Correspondence and requests for materials should be addressed to W.A.C. (e-mail: wcatt@u.washington.edu).

Calmodulin supports both inactivation and facilitation of L-type calcium channels

Roger D. Zühlke^{*†}, Geoffrey S. Pitt^{‡§}, Karl Deisseroth[§], Richard W. Tsien[§] & Harald Reuter^{*}

^{*} Department of Pharmacology, University of Bern, 3010 Bern, Switzerland

[‡] Division of Cardiovascular Medicine and [§] Department of Molecular and Cellular Physiology, Stanford University Medical School, Stanford, California 94305-5426, USA

[†] These authors contributed equally to this work.

L-type Ca^{2+} channels support Ca^{2+} entry into cells, which triggers cardiac contraction¹, controls hormone secretion from endocrine cells² and initiates transcriptional events that support learning and memory³. These channels are examples of molecular signal-transduction units that regulate themselves through their own activity. Among the many types of voltage-gated Ca^{2+} channel, L-type Ca^{2+} channels particularly display inactivation and facilitation, both of which are closely linked to the earlier entry of Ca^{2+} ions^{4–10}. Both forms of autoregulation have a significant impact on the amount of Ca^{2+} that enters the cell during repetitive activity, with major consequences downstream. Despite extensive biophysical analysis⁹, the molecular basis of autoregulation remains unclear, although a putative Ca^{2+} -binding EF-hand motif^{11,12} and a nearby consensus calmodulin-binding isoleucine-glutamine ('IQ') motif^{13,14} in the carboxy terminus of the α_{1C} channel subunit have been implicated^{12,14–16}. Here we show that calmodulin is a critical Ca^{2+} sensor for both inactivation and facilitation, and that the nature of the modulatory effect depends on residues within the IQ motif important for calmodulin binding. Replacement of the native isoleucine by alanine removed Ca^{2+} -dependent inactivation and unmasked a strong facilitation; conversion of the same residue to glutamate eliminated both forms of autoregulation. These results indicate that the same calmodulin molecule may act as a Ca^{2+} sensor for both positive and negative modulation.

Deletion of the first eight amino acids of the IQ motif in the C-terminal tail of the Ca^{2+} -channel subunit α_{1C} (Fig. 1a) eliminates Ca^{2+} -dependent inactivation¹⁴, but there is no information about any interaction with calmodulin (CaM) or its functional role in Ca^{2+} -dependent inactivation or facilitation. Mutation of the isoleucine residue Ile1624 to alanine, valine or glutamate (77I/A, 77I/V, 77I/E; Fig. 1b) revealed that this amino acid has a critical role in Ca^{2+} -dependent inactivation and facilitation. Complementary RNAs (cRNAs) for wild-type α_{1C} (77WT)^{14,17} or its mutants were injected into *Xenopus* oocytes, together with cRNAs for the auxiliary subunits $\alpha_2\delta$ and β_1 . Both Ba^{2+} and Ca^{2+} currents (I_{Ba} and I_{Ca}) through the expressed channels were recorded in every oocyte (representative traces are shown in Fig. 1c). As reported¹⁴, channel 77WT shows prominent Ca^{2+} -dependent inactivation. In

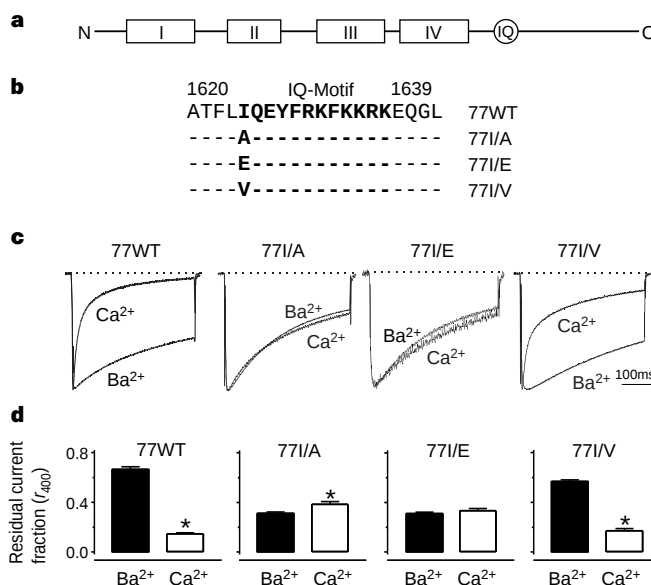


Figure 1 Point mutations in the IQ motif of 77WT affect Ca^{2+} -dependent inactivation. **a**, 77WT with four transmembrane domains (I–IV) and a consensus CaM-binding IQ motif 148 amino acids downstream of transmembrane segment S6 in domain IV. **b**, Sequence alignment of amino acids 1,620–1,639 of 77WT (IQ motif in bold) and its variants with the respective point mutations at position 1,624. **c**, Representative currents recorded from 77WT and its mutants during test-pulses of V_h from -90 to $+20$ mV. I_{Ca} was scaled to peak I_{Ba} . **d**, Residual fractions of peak currents remaining at the end of a 400-ms test-pulse (r_{400}), plotted for I_{Ba} and I_{Ca} supported by 77WT, 77I/A, 77I/E and 77I/V ($n = 8$ –20 each). Asterisk, $P < 0.001$ vs. I_{Ba} , paired t -test.

contrast, channel mutants 77I/A and 77I/E did not show this type of inactivation. In fact, the decay of I_{Ca} through 77I/A was sometimes even slightly slower than that of I_{Ba} . However, channel 77I/V showed Ca^{2+} -dependent inactivation similar to that of 77WT (Fig. 1c). Data were pooled using the ratio of the current values at 400 ms and at the initial peak (r_{400} , Fig. 1d) as an index of inactivation. The residual current fraction (r_{400}) for Ca^{2+} was significantly smaller than that for Ba^{2+} for both 77WT and 77I/V, but not for 77I/A and 77I/E.

Unexpectedly, mutations at amino-acid residue position 1,624 also influenced Ca^{2+} -dependent facilitation during trains of voltage pulses (Fig. 2). Ca^{2+} currents supported by 77I/A increased markedly with repeated depolarizations to $+20$ mV, whereas Ba^{2+} currents increased very little (Fig. 2a, b). This can be seen in records of I_{Ca} and I_{Ba} (Fig. 2a), scaled according to their peak amplitudes at the basal frequency of stimulation (0.1 Hz) to allow for intrinsic differences in open channel flux¹⁸. Peak I_{Ca} showed increasingly strong facilitation as pulse frequency was raised, reaching 72% above baseline at 3.3 Hz, but peak I_{Ba} increased only marginally at higher frequencies. The facilitation was an incremental phenomenon, as illustrated with trains of 40 pulses (Fig. 2b). In contrast to 77I/A, 77WT and mutant channels 77I/E and 77I/V exhibited little or no overt potentiation at any frequency between 0.5 and 3.3 Hz, with either Ca^{2+} or Ba^{2+} as charge carriers (Fig. 2c).

Might the strong facilitation observed with mutant 77I/A be a general channel property that is sometimes masked by Ca^{2+} -dependent inactivation? We used a two-pulse protocol to look in more detail at the kinetics of recovery from the effects of an initial conditioning pulse (Fig. 3). The pattern of differences between I_{Ba} and I_{Ca} varied strikingly, depending on the amino acid at position 1,624. In 77WT (and in 77I/V), the Ca^{2+} -dependence of inactivation was manifested by a lesser degree of recovery with Ca^{2+} than with Ba^{2+} as charge carrier, seen over the first ~ 30 ms of recovery¹⁴. At later times, however, the differential reversed, with faster recovery

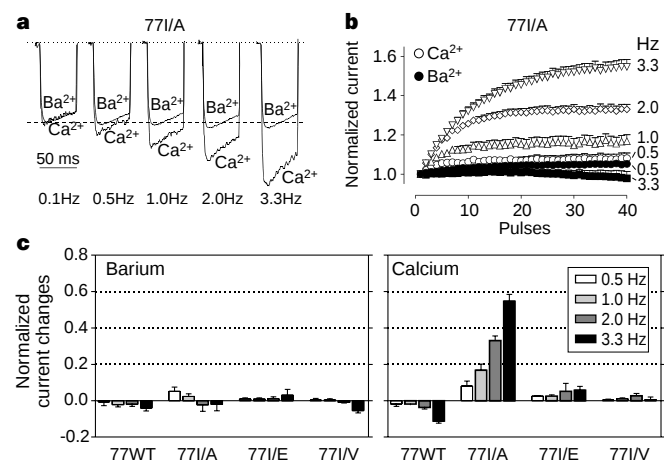


Figure 2 Frequency-dependent facilitation of I_{Ca} conducted by 77I/A. **a**, I_{Ba} and I_{Ca} current traces taken at the end of trains of 40 test pulses of V_h from -90 to $+20$ mV, at 0.1 to 3.3 Hz. I_{Ba} and I_{Ca} traces were normalized to their respective scaled amplitude at 0.1 Hz (dashed line). **b**, Peak I_{Ba} (filled symbols) and I_{Ca} (open symbols) during trains of 40 repetitive test-pulses at 0.1 – 3.3 Hz were normalized to the current amplitude at the beginning of each train. **c**, Changes in peak I_{Ba} (left panel) and I_{Ca} (right panel) conducted by 77WT, 77I/A, 77I/E and 77I/V ($n = 4$ – 12 each) at indicated stimulation frequencies. Note that frequency-dependent facilitation was present only in 77I/A.

for I_{Ca} . In contrast, with 77I/A, I_{Ca} recovered faster than I_{Ba} at all times, reaching a peak value 20% greater than that during the first pulse after a rest period of >200 ms. For 77I/E, recovery of I_{Ca} and I_{Ba} was rapid and indistinguishable at all recovery intervals. These results indicate that the amino acids at position 1,624 are critical for both early and late differences. The late differences can be interpreted as facilitation, which can be observed in combination with Ca^{2+} -dependent inactivation (77WT and 77I/V), in isolation (77I/A), or not at all (77I/E).

As isoleucine is the first amino acid in the putative CaM-binding IQ motif in 77WT and its mutation had such a profound effect on the channel's Ca^{2+} -sensitive feedback mechanisms, we tested whether binding of CaM to the IQ motif could be involved in the regulation of Ca^{2+} -dependent inactivation and facilitation.

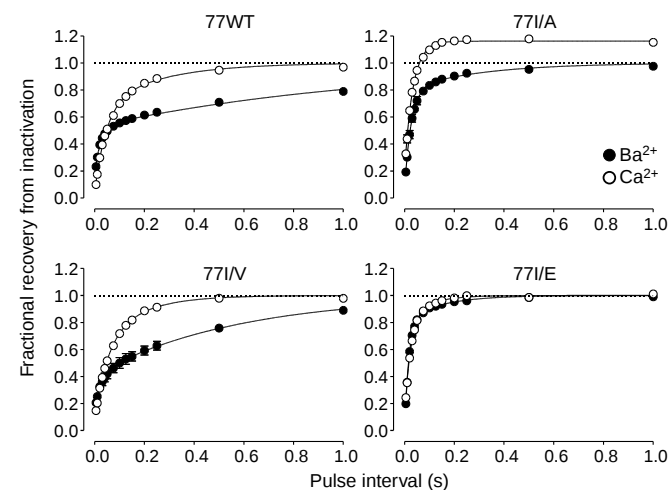


Figure 3 Repriming experiments reveal Ca^{2+} -dependent facilitation of 77WT and 77I/V. Recovery from inactivation of I_{Ba} and I_{Ca} was studied with a standard two-pulse protocol. During prepulses of V_h from -90 to $+20$ mV, currents were reduced to 15–20%, and subsequent test pulses applied at variable intervals after prepulses revealed the kinetics of recovery from inactivation. Fractional recovery of peak I_{Ba} and I_{Ca} for 77WT and its mutants are plotted against the interpulse interval. Solid lines represent averaged bi-exponential fits of 4–7 experiments.

Although previous experiments have failed to demonstrate any effect of the CaM inhibitor calmidazolium on Ca^{2+} -dependent inactivation^{9,14}, a role for CaM would not be ruled out by these experiments if CaM was protected from calmidazolium at basal Ca^{2+} concentrations¹⁹. Accordingly, we tested the effects of a CaM mutant (here called CaM(3–)) that contained alanine instead of aspartate in three out of the four Ca^{2+} -binding EF-hand motifs¹⁹. When CaM(WT) was co-expressed with Ca^{2+} -channel subunits in oocytes, there was little effect on the pattern of inactivation of I_{Ba} and I_{Ca} (Fig. 4) compared to controls lacking exogenous CaM (Fig. 1c, d). However, CaM(3–) strongly inhibited Ca^{2+} -dependent inactivation of channels 77WT and 77I/V (Fig. 4a, b) and also completely eliminated Ca^{2+} -dependent facilitation of I_{Ca} through channel 77I/A (Fig. 4c, d). There was no effect on facilitation after the injection of cRNA for CaM(WT) (Fig. 4c, d). These results demonstrate that CaM plays a pivotal role in both Ca^{2+} -dependent inactivation and facilitation.

Having found that amino-acid alterations within the conserved IQ motif cause significant changes in Ca^{2+} -dependent plasticity of channel function, and having demonstrated that CaM is involved, we investigated whether CaM binds to the channel motif in a channel-specific manner. We first examined whether CaM binds to the entire α_{1C} -cytoplasmic tail (Fig. 5a). A ^{35}S -labelled *in vitro*-translated α_{1C} -cytoplasmic tail did bind to immobilized CaM in the presence of saturating Ca^{2+} , but not in the absence of Ca^{2+} . Mutating Ile 1624 to glutamate, which extinguished the CaM-dependent inactivation and facilitation in the native channel, decreased CaM binding more than 5-fold (Fig. 5a). Ca^{2+} -dependent interactions were examined in experiments using a 20-amino-acid peptide encompassing the IQ region of 77WT (α_{1C} in Fig. 5b). We tested for possible binding between the peptide and CaM in gel-shift assays

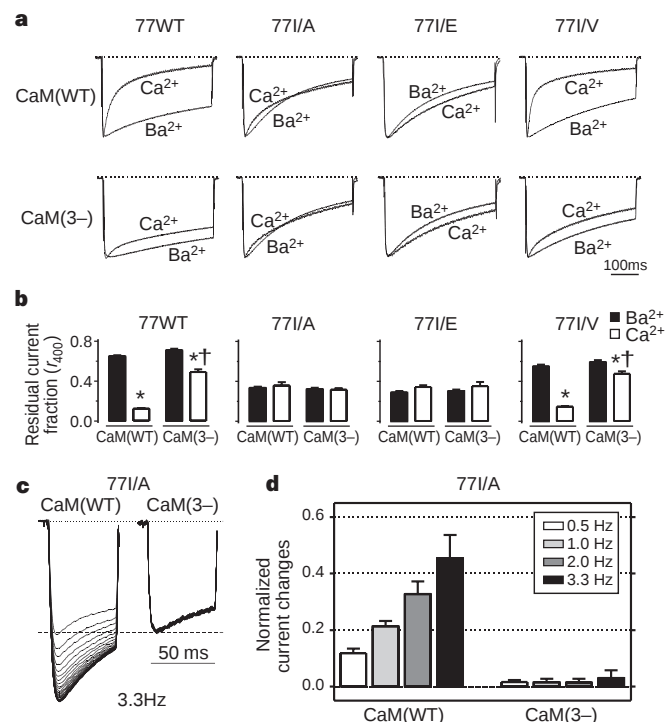


Figure 4 A mutant calmodulin, CaM(3–), inhibits Ca^{2+} -dependent inactivation and facilitation. **a**, I_{Ba} and scaled I_{Ca} traces recorded from oocytes expressing 77WT or its mutants together with CaM(WT) (upper panel) or CaM(3–) (lower panel). **b**, Differences between r_{400} values of I_{Ca} and I_{Ba} through 77WT and 77I/V ($n = 6$ – 16) were pronounced with CaM(WT), but reduced with CaM(3–); no significant differences in 77I/A and 77I/E ($n = 4$ – 10). **c**, Co-expression of 77I/A with CaM(3–), but not with CaM(WT), eliminated I_{Ca} facilitation with repetitive pulsing. **d**, Plot of changes at various frequencies. Asterisk, $P < 0.001$ vs. I_{Ba} ; dagger, $P < 0.001$ vs CaM(WT); both by ANOVA and Bonferroni tests.

in which mixtures of both were run on non-denaturing polyacrylamide gels (Fig. 5c). With 100 nM Ca^{2+} , which is comparable to the basal Ca^{2+} level in quiescent cells, inclusion of peptide reduced the mobility of CaM. The mobility shift saturated at a peptide:CaM ratio corresponding to a 1:1 complex. In contrast, no shift was found in the absence of Ca^{2+} (data not shown). For comparison, we also tested a peptide from the corresponding region of α_{1B} (N-type channel), which diverged at three positions among the IQ consensus residues (Fig. 5b). This α_{1B} -derived peptide did not alter CaM mobility at either 100 nM (Fig. 5c) or 1 mM free Ca^{2+} (data not shown).

An additional perspective on the CaM-peptide interaction was obtained in fluorescence experiments with dansyl-CaM²⁰. The dansyl-CaM emission spectrum displayed an IQ-peptide-dependent blue shift and enhancement, characteristic of a hydrophobic CaM-peptide interaction (Fig. 5d). The apparent K_m for the α_{1C} peptide was 30-fold lower with 2 mM Ca^{2+} than without Ca^{2+} , whereas the α_{1B} -derived peptide displayed only a weak interaction with dansyl-CaM and did not show a Ca^{2+} -dependent increase in

affinity (data not shown). The Ca^{2+} -dependence of the dansyl-CaM- α_{1C} peptide interaction (Fig. 5e) was detected as a steep increase in the fluorescence signal over submicromolar Ca^{2+} concentrations. The data were well-fitted by a simple model (see Methods) that allowed different interactions of dansyl-CaM with peptide depending on whether it was complexed with zero, two or four Ca^{2+} ions. Peptide-CaM (P-CaM) binding was favoured by a high-affinity P-CaM- Ca^{2+} interaction ($K_{D1} = 29$ nM), which would give rise to binding at Ca^{2+} levels in quiescent cells (Fig. 5c). However, the main increment in fluorescence arose from a further P-CaM- Ca^{2+} interaction of lower affinity ($K_{D2} = 515$ nM), presumably corresponding to CaM effector action.

Our experiments revealed some novel features of L-type-channel gating: first, CaM plays a pivotal role in both Ca^{2+} -dependent inactivation and facilitation; second, a putative IQ motif in the cytoplasmic tail of α_{1C} not only binds CaM but is also important for the opposing modulatory actions; and third, mutation of Ile 1624 to alanine inhibits Ca^{2+} -dependent inactivation and unmasks overt facilitation, whereas its conversion to glutamate eliminates both effects. The involvement of CaM in inactivation had previously been considered¹⁴, but was not substantiated because of the ineffectiveness of calmidazolium^{9,14}. This CaM inhibitor would only be expected to block Ca^{2+} -dependent inactivation if CaM acted as a free molecule, but CaM appears to be stably associated with the IQ motif at basal Ca^{2+} levels (Fig. 5c). Such an association may help account for the persistence of Ca^{2+} -dependent inactivation in excised patches²¹ and planar bilayers²². Overexpression of mutant CaM(3-) exerted a dominant negative effect on inactivation (Fig. 4), but this behaviour only became prominent after expression in oocytes over several days, time presumably needed for turnover and displacement of endogenous CaM. The C-terminal CaM-binding site strategically positions a cytoplasmic Ca^{2+} sensor close to the permeation pathway, favouring local feedback control of channel function. This parallels CaM's involvement in regulating SK channels¹⁹ (in which CaM affects gating by interacting with domains in analogous positions), NMDA channels²³, cyclic-nucleotide-gated cation channels²⁴ and Ca^{2+} pumps²⁵. The L-type Ca^{2+} channel is unusual in using CaM as a regulator of both inactivation and facilitation, which indicates that the same CaM molecule may act as a Ca^{2+} sensor for opposing biological effects. □

Methods

Molecular biology. Site-directed mutagenesis was done using the plasmid p77NB¹⁴ as template. A mutagenic oligonucleotide which introduced the degenerate codon GNA (where N is G, A, T or C) encoding amino-acid residues at position 1,624 was used in PCR reactions. An amplified 400-bp *Var911/RsaI*-restriction fragment containing the point mutations was subcloned by triple-fragment ligation directly into the plasmid pHLCC77²⁶, which encodes the full length 77WT-channel subunit. We confirmed the mutations by sequencing the entire 400-bp fragment. *In vitro* transcription of the wild-type and mutant-77 constructs and of the auxiliary constructs $\alpha_2\delta$ (ref. 27) and β_1 (ref. 28) was as described¹⁴. Transcription of CaM(WT) and CaM(3-)¹⁹ was performed with *PvuI*-linearized complementary DNA templates and Ambion's T7 mMessage mMachine Kit. The transcription reactions were spiked with 25 units of SP6 RNA polymerase. Isolation and handling of *Xenopus* oocytes followed standard procedures²⁹. For co-expression experiments of Ca^{2+} channels with CaM(WT) and CaM(3-), respectively, up to 200 fmol of CaM(WT) and CaM(3-) cRNAs were injected 2–3 days before the injection of Ca^{2+} -channel subunit cRNAs.

Electrophysiology. Whole-cell I_{Ba} and I_{Ca} were recorded in every oocyte by a standard two-electrode voltage clamp method¹⁴. To eliminate Cl^- currents through endogenous Ca^{2+} -activated channels, we injected 23–46 nl of 100 mM BAPTA (1,2-bis(2-aminophenoxy)ethane-N,N,N',N'-tetraacetic K_4 hydrate) solution into the oocytes 1–3 h before the recordings. During the recordings, oocytes were superfused with a solution containing (in mM): Ba(OH)₂ or Ca(NO₃)₂ 40, NaOH 50, KOH 1, HEPES 10 (adjusted to pH 7.4 with

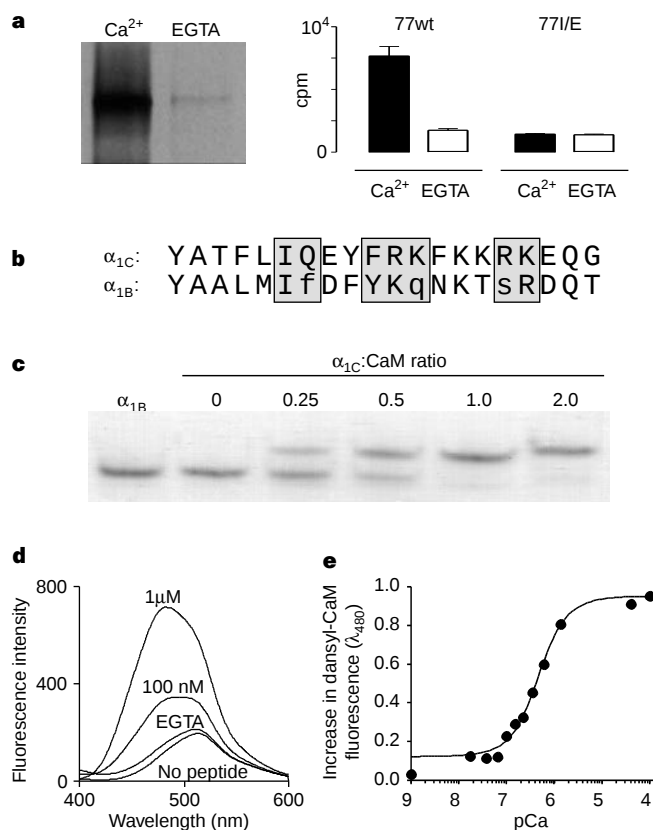


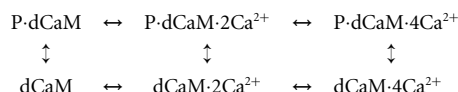
Figure 5 CaM interacts with the IQ motif of α_{1C} . **a**, Left, interaction between CaM and ³⁵S-labelled, *in vitro*-translated α_{1C} C-terminal fragment; right, comparison of wild-type and I/E C-terminal fragments with regard to CaM interaction. **b**, Peptide sequences of consensus IQ region of α_{1C} and corresponding region of α_{1B} . Divergence from consensus indicated in lower case. **c**, Gel mobility-shift assay demonstrating the interaction between CaM and the α_{1C} IQ peptide. CaM was incubated with IQ peptide at the indicated molar ratios in solution containing ~100 nM free [Ca^{2+}]. First lane shows lack of mobility shift with α_{1B} peptide (2:1 ratio over CaM). **d**, Emission fluorescence spectra of dansyl-CaM in the absence (No peptide) or presence of α_{1C} peptide with no added Ca^{2+} (EGTA), 100 nM and 1 μM added Ca^{2+} ; spectra have been corrected for background buffer fluorescence. **e**, Ca^{2+} -dependence of peak fluorescence due to interaction between dansyl-CaM (dCaM, 500 nM) and α_{1C} IQ peptide (500 nM). Data points show averages of two independent runs. Smooth curve was derived from a simplified model (Methods) in which dCaM may be complexed with zero, two or four Ca^{2+} ions, either with or without bound IQ peptide. Least-squares fit yielded values of $K_{D1} = 29$ nM, $K_{D2} = 515$ nM.

methanesulphonic acid). Currents were filtered at 0.5 Hz and sampled at 2 Hz, and leak subtraction was performed using an offline P/4 protocol. Data points are mean $\pm$ s.e.m.

CaM- α_{1C} interaction. A PCR fragment containing nucleotides 4,710–6,707 was generated from rabbit α_{1C} cDNA and cloned into pCDNA3. Wild-type and I/E sequences were studied. 35 S-labelled *in vitro* translated peptide was generated in rabbit reticulocyte lysates, separated from unincorporated 35 S by passage over a G50 column, and incubated for 4 h at 4°C with CaM immobilized on agarose (Calbiochem) in buffer solutions containing 150 mM NaCl, 50 mM Tris–HCl (pH 7.2) plus 0.1% Triton X-100, with either 1 mM CaCl_2 or 2 mM EGTA. The CaM-agarose was then washed extensively with the same buffer and the labelled peptide was eluted in SDS, and either quantified by scintillation counting or separated on SDS–PAGE and identified by autoradiography.

Gel mobility-shift assay. CaM (1.5 nmol) was incubated with α_{1C} IQ peptide for 1 h at 22°C in 50 mM Tris, pH 7.4, 2 μM CaCl_2 and 2.5 μM EGTA (calculated free $[\text{Ca}^{2+}] \approx 100$ nM). We subjected the sample to non-denaturing polyacrylamide-gel electrophoresis in the presence of 100 nM free Ca^{2+} before visualizing the CaM complex by staining with Coomassie blue.

Dansyl–CaM studies. Conjugation of highly purified CaM with dansyl (5-dimethyl-aminonaphthalene-1-sulphonyl) chloride (Molecular Probes) was performed for 1 h at room temperature. We purified the product with G-25 Sephadex columns, and carried out binding studies in 150 mM NaCl and 50 mM Tris (pH 7.5), with or without Ca^{2+} as indicated. Fluorescence was monitored with a Perkin-Elmer LS50B luminescence spectrometer with 340 nm excitation. We studied the Ca^{2+} -dependence using a series of Ca^{2+} buffers (Molecular Probes) containing 100 mM KCl, 10 mM MOPS (pH 7.2), 1 mM free Mg^{2+} , and Ca^{2+} -EGTA concentrations adjusted to give the indicated free Ca^{2+} . Mg^{2+} could not functionally substitute for Ca^{2+} in controlling peptide binding (not shown). Peptides were obtained from Research Genetics. The smooth curve describing the fractional increase in fluorescence was derived from a simplified model in which dansyl–CaM (dCaM) may be complexed with 0, 2 or 4 Ca^{2+} ions, either without or with bound IQ peptide (P).



We assume that occupancy of the dCaM-2 Ca^{2+} state is insignificant, because dCaM-2 Ca^{2+} would either bind very rapidly to peptide or take on two additional Ca^{2+} ions. Peptide dissociation constants for P-dCaM (10.8 μM) and P-dCaM-4 Ca^{2+} (304 nM) were experimentally determined in 0 or 2 mM free Ca^{2+} . $\text{P-dCaM} \leftrightarrow \text{P-dCaM-2Ca}^{2+}$ and $\text{P-dCaM-2Ca}^{2+} \leftrightarrow \text{P-dCaM-4Ca}^{2+}$ reactions were described as 1:1 reactions for simplicity, using $K_{1/2}$ values (K_{D1} , K_{D2}) adjusted by least-squares criterion to fit the experimental data. Calculation of aggregate fluorescence took into account estimates of the relative fluorescence of individual states, dCaM (0.05), P-dCaM (and, by assumption, P-dCaM-2 Ca^{2+}) (0.229), P-dCaM-4 Ca^{2+} (0.657) and dCaM-4 Ca^{2+} (0.123).

Received 25 January; accepted 24 March 1999.

- Bers, D. M. *Excitation–Contraction Coupling and Cardiac Contractile Force* (Kluwer Academic, Dordrecht, 1991).
- Artalejo, C. R., Adams, M. E. & Fox, A. P. Three types of Ca^{2+} channels trigger secretion with different efficacies in chromaffin cells. *Nature* **367**, 72–76 (1994).
- Murphy, T. H., Worley, P. F. & Baraban, J. M. L-type voltage-sensitive calcium channels mediate synaptic activation of immediate early genes. *Neuron* **7**, 625–635 (1991).
- Eckert, R. & Chad, J. E. Inactivation of Ca^{2+} channels. *Prog. Biophys. Mol. Biol.* **44**, 215–267 (1984).
- Noble, S. & Shimoni, Y. The calcium and frequency dependence of the slow inward current 'staircase' in frog atrium. *J. Physiol.* **310**, 57–75 (1981).
- Gurney, A. M., Charnet, P., Pye, J. M. & Nargeot, J. Augmentation of cardiac calcium current by flash photolysis of intracellular caged- Ca^{2+} molecules. *Nature* **341**, 65–68 (1989).
- Zygmunt, A. C. & Maylie, J. Stimulation-dependent facilitation of the high threshold calcium current in guinea-pig ventricular myocytes. *J. Physiol.* **428**, 653–671 (1990).
- McDonald, T. F., Pelzer, S., Trautwein, W. & Pelzer, D. J. Regulation and modulation of calcium channels in cardiac, skeletal, and smooth muscle cells. *Physiol. Rev.* **74**, 365–507 (1994).
- Imredy, J. P. & Yue, D. T. Mechanism of Ca^{2+} -sensitive inactivation of L-type Ca^{2+} channels. *Neuron* **12**, 1301–1318 (1994).
- Soldatov, N. M., Zühlke, R. D., Bouron, A. & Reuter, H. Molecular structures involved in L-type calcium channel inactivation. Role of the carboxyl-terminal region encoded by exons 40–42 in α_{1C} subunit in the kinetics and Ca^{2+} -dependence of inactivation. *J. Biol. Chem.* **272**, 3560–3566 (1997).
- Babitch, J. Channel hands. *Nature* **346**, 321–322 (1990).
- De Leon, M. et al. Essential Ca^{2+} -binding motif for Ca^{2+} -sensitive inactivation of L-type Ca^{2+} channels. *Science* **270**, 1502–1506 (1995).
- Rhoads, A. R. & Friedberg, F. Sequence motifs for calmodulin recognition. *FASEB J.* **11**, 331–340 (1997).
- Zühlke, R. D. & Reuter, H. Ca^{2+} -sensitive inactivation of L-type Ca^{2+} channels depends on multiple cytoplasmic amino acid sequences of the α_{1C} subunit. *Proc. Natl Acad. Sci. USA* **95**, 3287–3294 (1998).
- Zhou, J. M. et al. Feedback inhibition of Ca^{2+} channels by Ca^{2+} depends on a short sequence of the C

- terminus that does not include the Ca^{2+} -binding function of a motif with similarity to Ca^{2+} -binding domains. *Proc. Natl Acad. Sci. USA* **94**, 2301–2305 (1997).
- Bernatchez, G., Talwar, D. & Parent, L. Mutations in the EF-hand motif impair the inactivation of barium currents of the cardiac α_{1C} channel. *Biophys. J.* **75**, 1727–1739 (1998).
- Soldatov, N. M. Molecular diversity of L-type Ca^{2+} channel transcripts in human fibroblasts. *Proc. Natl Acad. Sci. USA* **89**, 4628–4632 (1992).
- Hess, P., Lansman, J. B. & Tsien, R. W. Calcium channel selectivity for divalent and monovalent cations. Voltage and concentration dependence of single channel current in ventricular heart cells. *J. Gen. Physiol.* **88**, 293–319 (1986).
- Xia, X. M. et al. Mechanism of calcium gating in small-conductance calcium-activated potassium channels. *Nature* **395**, 503–507 (1998).
- Kincaid, R. L., Vaughan, M., Osborne, J. C. Jr & Tkachuk, V. A. Ca^{2+} -dependent interaction of 5-dimethylaminonaphthalene-1-sulfonyl-calmodulin with cyclic nucleotide phosphodiesterase, calcineurin, and troponin I. *J. Biol. Chem.* **257**, 10638–10643 (1982).
- Höfer, G. F. et al. Intracellular Ca^{2+} inactivates L-type Ca^{2+} channels with a Hill coefficient of ~ 1 and an inhibition constant of $\sim 4 \mu\text{M}$ by reducing channel's open probability. *Biophys. J.* **73**, 1857–1865 (1997).
- Haack, J. A. & Rosenberg, R. L. Calcium-dependent inactivation of L-type calcium channels in planar lipid bilayers. *Biophys. J.* **66**, 1051–1060 (1994).
- Zhang, S., Ehlers, M. D., Bernhardt, J. P., Su, C. T. & Hagan, R. L. Calmodulin mediates calcium-dependent inactivation of N-methyl-D-aspartate receptors. *Neuron* **21**, 443–453 (1998).
- Liu, M., Chen, T., Ahmed, B., Li, J. & Yau, K. W. Calcium-calmodulin modulation of the olfactory cyclic nucleotide-gated cation channel. *Science* **266**, 1348–1354 (1994).
- Vorherr, T. et al. Interaction of calmodulin with the calmodulin binding domain of the plasma membrane Ca^{2+} pump. *Biochemistry* **29**, 355–365 (1990).
- Soldatov, N. M., Bouron, A. & Reuter, H. Different voltage-dependent inhibition by dihydropyridines of human Ca^{2+} channel splice variants. *J. Biol. Chem.* **270**, 10540–10543 (1995).
- Singer, D. et al. The role of the subunits in the function of the calcium channel. *Science* **253**, 1553–1557 (1991).
- Ruth, P. et al. Primary structure of the β subunit of the DHP-sensitive calcium channel from skeletal muscle. *Science* **245**, 1115–1118 (1989).
- Zühlke, R. D., Bouron, A., Soldatov, N. M. & Reuter, H. Ca^{2+} channel sensitivity towards the blocker isradipine is affected by alternative splicing of the human α_{1C} subunit gene. *FEBS Lett.* **427**, 220–224 (1998).

Acknowledgements. We thank J. Adelman and J. Maylie for providing CaM(WT) and CaM(–3), C. Klee and M. Ikura for helpful discussions, and H. van Hees for technical assistance. This work was supported by the Swiss National Science Foundation (H.R.), a Pfizer postdoctoral fellowship (G.S.P.), an MSTP training grant (K.D.), and grants from NIH (G.S.P.), NINDS, NIMH, the McKnight Foundation and the Mathers Charitable Trust (R.W.T.).

Correspondence and requests for materials should be addressed to H.R. (e-mail: harald.reuter@pki.unibe.ch).

A cytosolic catalase is needed to extend adult lifespan in *C. elegans daf-C* and *clk-1* mutants

James Taub[†], Joe F. Lau^{*}, Charles Ma^{*}, Jang Hee Hahn[†], Rafaz Hoque^{*}, Jonathan Rothblatt[†] & Martin Chalfie^{*}

^{*} Department of Biological Sciences, Columbia University, 1212 Amsterdam Avenue, New York, New York 10027, USA

[†] Department of Biological Sciences, 6044 Gilman Laboratory, Dartmouth College, Hanover, New Hampshire 03755, USA

The dauer larva is an alternative larval stage in *Caenorhabditis elegans* which allows animals to survive through periods of low food availability. Well-fed worms live for about three weeks, but dauer larvae can live for at least two months without affecting post-dauer lifespan¹. Mutations in *daf-2* and *age-1*, which produce a dauer constitutive (Daf-C) phenotype, and in *clk-1*, which are believed to slow metabolism, markedly increase adult lifespan². Here we show that a *ctl-1* mutation reduces adult lifespan in otherwise wild-type animals and eliminates the *daf-c* and *clk-1*-mediated extension of adult lifespan. *ctl-1* encodes an unusual cytosolic catalase; a second gene, *ctl-2*, encodes a peroxisomal catalase. *ctl-1* messenger RNA is increased in dauer larvae and adults with the *daf-c* mutations. We suggest that the *ctl-1* catalase is needed during periods of starvation, as in the dauer larva, and that its misexpression in *daf-c* and *clk-1* adults extends lifespan. Cytosolic catalase may have evolved to protect nematodes from

[‡] Present addresses: Department of Anatomy, College of Medicine, Kangwon National University, Chuncheon, 200-701, Korea (J.H.H.); Center for Applied Genomics, Hoechst Marion Roussel, Fraunhoferstr. 22, 82152 Martinsried, Germany (J.R.).

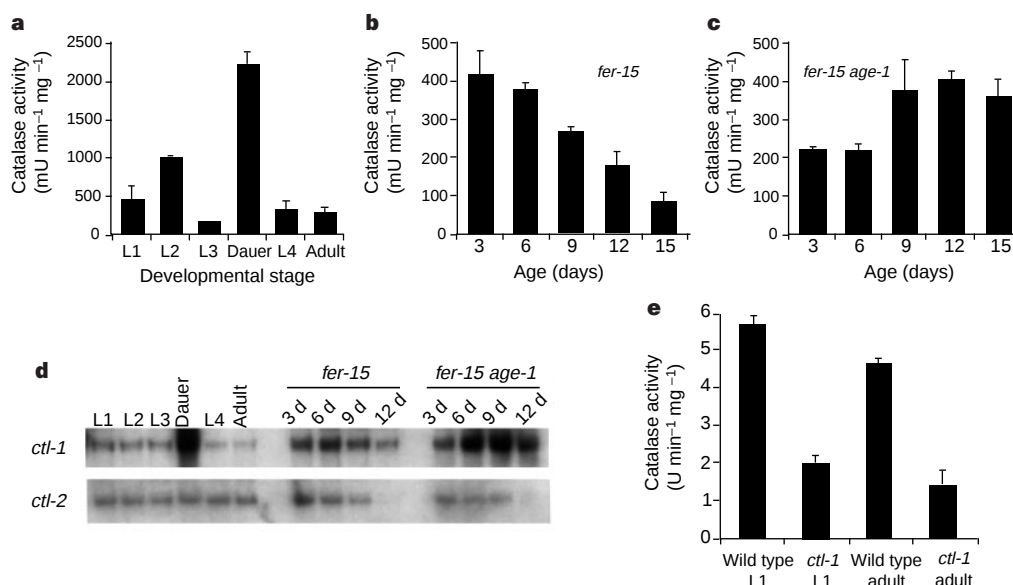


Figure 1 Catalase expression in wild-type and mutant animals. **a**, Total catalase activity (cytosolic and peroxisomal) in wild-type animals is highest in the L2 larval stage and declines slightly during normal development. Catalase activity is fivefold higher in *daf-2* dauer larvae than in other larval stages. Each value here and in other figures is the mean $\pm$ s.d. of three independent experiments. **b**, Catalase activity in ageing *fer-15* (*b26ts*) animals at 25°C. The steady decline in catalase activity as these animals age mirrors the appearance of signs of ageing, such as decreased movement and feeding activity. **c**, Catalase activity in *age-1* (*hx546*) *fer-15* (*b26ts*) animals increases with age. **d**, Abundance of *ctl-1* and

ctl-2 mRNAs during development and ageing of *C. elegans*. The level of *ctl-1* mRNA is highest in dauer larvae and older *age-1* adults. The abundance of *ctl-2* mRNA remains constant throughout early development and declines as wild-type adults age. Lanes labelled with the larval stages (L1–L4) and adult contain 5 μ g wild-type RNA. The RNA from dauer larvae was obtained from *daf-2* (*e1370*) animals grown at 25°C. **e**, Catalase activity is reduced by more than 50% in the *ctl-1* mutant compared to the N2 wild-type strain. Activity is reduced in both larval (L1) and adult stages.

oxidative damage produced during prolonged dormancy before reproductive maturity, or it may represent a general mechanism for permitting organisms to cope with the metabolic changes that accompany starvation.

Total catalase activity in wild-type (N2) worms is highest in the early larval stages (L1 and L2) and decreases as the worms mature (Fig. 1a). Catalase activity is more than fivefold higher in the *daf-2* dauer larva than in normal L3 animals (see also ref. 3). These results indicate that catalase is important in dauer larvae. Catalase could provide protection against hydrogen peroxide that is generated as a by-product of peroxisomal fatty-acid β -oxidation or offer a more general antioxidant defence.

Catalase activity decreases as *C. elegans* adults age, but the decrease is minimized in *age-1* mutants (Fig. 1b, c). Animals with *age-1* mutations displayed increased catalase activity and locomotor activity even very late in life, whereas control animals showed an earlier decline in catalase activity that mirrored the appearance of signs of ageing, such as decreased movement and feeding activity.

Using a partial *C. elegans* complementary DNA, cm20b12, identified as a catalase cDNA by the *C. elegans* Genome Project⁴, we probed a *C. elegans* cDNA library and recovered 19 cDNAs in two classes, *ctl-1* and *ctl-2*. The *ctl-1* DNA (GenBank accession no. U55384) differs from cm20b12, but sequences in the *ctl-2* cDNA matched both it and a complete cDNA sequence submitted to the EMBL database by K.J. Henkle-Duehrsen (accession no. X82175). *ctl-1* and *ctl-2* and the catalases they encode share extensive regions of identity at the nucleotide and amino-acid levels. The CTL-1 and CTL-2 catalases are 82% identical; the region of greatest divergence is the 50 amino acids at the carboxy terminal. Although the first 270 nucleotides of the *ctl-1* and *ctl-2* cDNAs are the same, the mRNAs originate from tandem genes on chromosome II.

Unlike the CTL-2 protein, which contains a conservative variant of the canonical type I peroxisomal targeting signal⁵ at its C terminal (SHI), CTL-1 lacks a peroxisomal targeting signal. To determine the subcellular localization of CTL-1 and CTL-2 in *C. elegans*, we

expressed gene fusions in which the green fluorescent protein gene (*gfp*) was fused to the 5' end of the entire coding region of *ctl-1* or *ctl-2*. As expected, worms expressing the *gfp::ctl-2* fusion from a heat-shock promoter displayed a punctate pattern of GFP expression typical of localization to peroxisomes, whereas animals expressing the *gfp::ctl-1* fusion displayed diffuse fluorescence that is consistent with cytosolic localization (Fig. 2). *C. elegans* thus has two different transcripts encoding catalases with different subcellular distributions. Because *C. elegans* does not have detectable glutathione peroxidase activity³, CTL-1 may fulfil the need for a cytosolic hydrogen peroxide scavenger.

The amount of *ctl-2* mRNA remains constant in wild-type animals as they develop to adulthood and does not change substantially in *daf-2* dauer larvae compared to wild-type L3 worms (Fig. 1d). With abundant food, the amount of *ctl-1* mRNA decreases slightly as the worms mature. This decrease in *ctl-1* mRNA mirrors the decrease in catalase activity measured in developing animals. In contrast to *ctl-2*, *ctl-1* mRNA is markedly increased in *daf-2* dauer larvae. These results suggest that *ctl-1* is responsible for the increase in catalase activity detected in dauer larvae. CTL-1 may act as a general scavenger of hydrogen peroxide in the cytosol, or it may protect cells from the effects of an unknown, dauer-specific

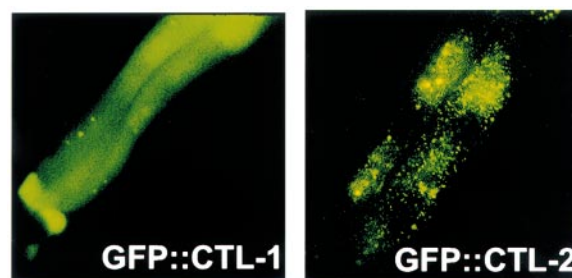


Figure 2 Heat-shock driven expression of *gfp::ctl-1* and *gfp::ctl-2* in the *C. elegans* intestine.

metabolic process. Until the metabolic processes occurring in the dauer larvae are defined, the role of CTL-1 in dauer metabolism will be unclear.

The patterns of *ctl-2* expression in *age-1* animals and their controls were identical: *ctl-2* mRNA declined as the animals aged. In contrast, *ctl-1* mRNA actually increased for 9 days and remained high 12 days after hatching. The abundance of *ctl-1* mRNA seen in *daf-2* dauer larvae and in old *age-1* animals suggests that *ctl-1* is needed for lifespan extension of both dauer larvae and *age-1* mutants. Moreover, these results indicate that *ctl-1*, and not *ctl-2*, is a downstream target of *daf-2* and *age-1*.

Using a subtractive cDNA screen (C.M. and M.C., unpublished data), we found a catalase cDNA whose mRNA was reduced at least 10-fold in the strain TU1061 (data not shown). The defect was a deletion of a single G in codon 42 (arginine) of *ctl-1*, resulting in a stop codon at codon 51. The *ctl-1(u800)* was crossed away from the other mutations in TU1061, and is maintained in strain TU2463. As expected, the total catalase activity of the resulting mutant animals was reduced by more than 50%, both as first-stage (L1) larvae and as adults, when compared to wild-type animals (Fig. 1e). The mean lifespan of the mutant animals was 77% of that of wild-type animals at 20°C (Table 1; Fig. 3a). Moreover, dauer survival was similarly reduced by the *ctl-1* mutation (Fig. 3b).

The *ctl-1(u800)* mutation causes a progeric phenotype. Although our criteria for death are that the animals fail to move in response to a prod from a platinum wire and exhibit no pharyngeal pumping, older wild-type and mutant animals are found that fail only one of these tests. *ctl-1* animals show these defects before wild-type animals. Another measure of ageing in *C. elegans* is the accumula-

Table 1 Effect of *ctl-1(u800)* on *C. elegans* lifespan

Genotype	50% survival*
Wild type	17.1 ± 0.3 (5)
<i>ctl-1</i>	13.0 ± 0.5 (5)
<i>daf-2</i>	22.5 ± 0.5 (3)
<i>ctl-1; daf-2</i>	13.2 ± 0.5 (3)
<i>age-1</i>	23.2 ± 1.4 (3)
<i>age-1 ctl-1</i>	13.1 ± 0.3 (3)
<i>daf-2; daf-12</i>	33.7 ± 1.5 (3)
<i>ctl-1; daf-2; daf-12</i>	13.7 ± 0.5 (4)
<i>clk-1</i>	20.5 ± 0.6 (3)
<i>ctl-1; clk-1</i>	14.7 ± 0.2 (3)
<i>daf-2 clk-1</i>	28.7 ± 0.6 (3)
<i>ctl-1; daf-2 clk-1</i>	16.2 ± 1.0 (3)

* Mean lifespan (days after hatching) ± s.d. The value in parentheses is the number of independent determinations of the mean lifespan for each strain.

tion of fluorescent material⁶. *ctl-1* mutants contain much more fluorescent material at 15 days after hatching than do wild-type animals (Fig. 4). These observations suggest that *ctl-1* is an important determinant of normal ageing in *C. elegans*. However, we have not seen any effect of transgenic expression of *ctl-1* on lifespan (even with co-expression of the superoxide dismutase gene *sod-1*).

The *ctl-1(u800)* mutation is epistatic to *daf-c* mutations with regard to their extension of lifespan, but not their effects on dauer formation. Several interacting pathways control the entry into and exit from the dauer stage¹. One pathway contains the *daf-c* genes *daf-2*, encoding an insulin receptor-like protein⁷, and *age-1* (also called *daf-23*), encoding a subunit of a phosphatidylinositol-3-kinase⁸. Mutations in these genes extend lifespan². With regard to dauer formation and lifespan extension, *daf-2* and *age-1* are

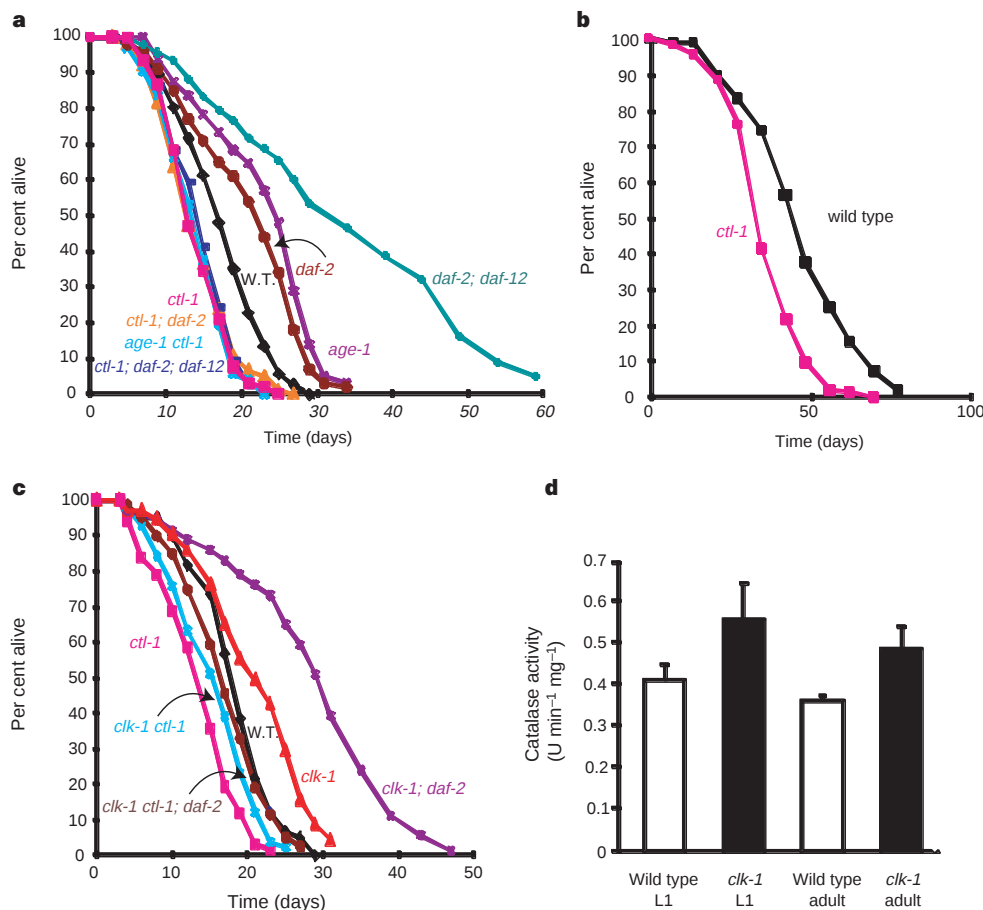


Figure 3 Effect of *ctl-1* activity on lifespan. **a**, The *ctl-1(u800)* mutation shortens lifespan compared to wild-type and is epistatic to the lifespan extension seen in *daf-2*, *age-1* and *daf-2; daf-12* animals. Data are shown from one experiment in Table 1 (other experiments produced equivalent data). **b**, *ctl-1(u800)* reduces

lifespan extension seen in *clk-1* and *clk-1; daf-2* animals. Data are shown from representative experiments summarized in Table 1. **d**, Catalase activity is increased in *clk-1* L3 larvae and young adults compared to wild-type.

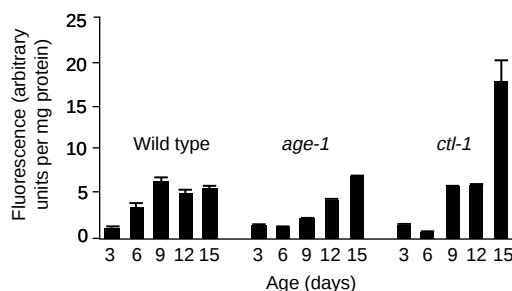


Figure 4 Autofluorescent material accumulates as animals age. The amount of fluorescent material is markedly higher in 15-day-old *ctl-1* animals as compared to 15-day-old wild-type or *age-1* animals. Error bars represent the s.d. of three independent samples.

genetically upstream of *daf-16*, a gene that encodes a forkhead-type transcription factor^{1,9,10}. An even greater extension of lifespan is seen in some *daf-2*;*daf-12* double mutants¹¹. Double and triple mutants containing the *daf-2*, *age-1*, and *daf-2*;*daf-12* mutations and *ctl-1* (*u800*) do not show any lifespan extension; they live no longer than animals possessing the *ctl-1* defect alone (Table 1; Fig. 3a). This epistasis suggests that *ctl-1* is downstream of these genes and that it is necessary for the lifespan extension seen in these mutants. Epistasis by *ctl-1* (*u800*) is only seen for lifespan extension; *ctl-1* mutants can form dauers and *ctl-1*;*daf-2* mutants show the temperature-sensitive DAF-C phenotype. We propose that the expression of *ctl-1* may be directly regulated by *daf-16* and that the inappropriate upregulation of *daf-16* in *daf-2* and *age-1* adults leads to misexpression of *ctl-1* and other cell-protective genes. This misexpression extends lifespan.

It is not clear whether the dauer-formation pathway that includes *daf-2*, *age-1* and *daf-16* affects the lifespan of wild-type adults, in part through the control of *ctl-1*: the data are ambiguous. If the pathway culminating in *daf-16* controlled the normal lifespan of *C. elegans*, loss of *daf-16* activity should reduce lifespan. This reduction has not been seen (except in one of three experiments in ref. 11) in *daf-16*(*m26*) animals, the only mutant animals for which results have been published. However, as the *m26* allele contains a splice-site mutation at the end of exon 2 (refs 9, 10), it may not produce a null phenotype. In contrast, the *daf-16*(*m26*) mutation does lower the total catalase activity slightly (data not shown), so a more complete loss of function may affect both catalase level and lifespan.

The *clk-1*(*e2519*) mutation also extends lifespan in *C. elegans*¹². It has been proposed¹² that the *clk-1* mutation extends lifespan by slowing metabolism. This may be true for larval animals, but our data are more consistent with the hypothesis that the *clk-1* mutation, directly or indirectly, induces *ctl-1* (and perhaps other antioxidant enzymes) which, in turn, extends lifespan by protecting cells from oxidative damage. The *ctl-1*(*u800*) mutation reduces but does not eliminate the lifespan extension seen in *clk-1*(*e2519*) and *daf-2*(*e1370*);*clk-1*(*e2519*) animals (Fig. 3c). These data demonstrate that the *ctl-1* mutation does not simply produce a time-dependent death of the animals. However, the difference in lifespan between *ctl-1* and *ctl-1*;*clk-1* animals is eliminated if only adult lifespan is considered. The time from hatching to adulthood at 20 °C for both *clk-1* and *ctl-1*;*clk-1* animals is between 5 and 6 days; that of wild-type and *ctl-1* animals is about 3 days. Thus, the *ctl-1* mutation appears to eliminate the extension of adult lifespan produced by the *clk-1* mutation, but does not affect the delay in larval development. To test whether catalase activity is important for the extension of adult lifespan by the *clk-1* mutation, we measured this activity in wild-type and *clk-1* animals at equivalent ages (3 days; data not shown) and at equivalent developmental stages (third-stage larvae and young adults; Fig. 3d). In all experiments, the *clk-1* animals had more catalase activity.

Our experiments with *ctl-1* suggest that the cytoplasmic catalase it encodes and oxidative stress defence are important determinants of lifespan in *C. elegans*. It has been proposed that oxidative damage to cells may be a major cause of cellular and organismal senescence¹³. In *Drosophila melanogaster*, systemic overexpression of Cu/Zn SOD and catalase, but not of either alone, increased mean adult lifespan by 33% (ref. 14 and references therein). These results indicate not only that control of reactive oxygen species is an important determinant of longevity, but also the need to balance SOD and catalase activities for the control of oxidative stress (see also ref. 15). This balance may be different for different cells, because overexpression of human SOD1 without additional catalase in *Drosophila* motor neurons also extends lifespan¹⁶.

We believe that *C. elegans* uses CTL-1 to cope with oxidative stress in the cytosol, where it may work with a superoxide dismutase. Counteracting cytosolic oxidative stress may allow animals to survive long periods of pre-reproductive dormancy as dauer larvae. One explanation for the effect of the *daf-2* and *age-1* mutations on ageing is that misexpression of the *daf-16* pathway, and particularly its induction of *ctl-1* expression, in late life causes the lifespan extension seen in the mutants. All somatic cells in *C. elegans* are postmitotic, so our results do not address the role of antioxidant enzymes in actively dividing tissues.

Many models for the evolution of senescence suggest that a trade-off is made between reproduction and life extension (for example, ref. 17). In such models, gene activity that extends lifespan, especially of the period after reproduction, would be selected against, because long-lived individuals would compete against their own progeny for necessary resources. Selection for such a gene is difficult to envision, as the affected individuals are post-reproductive. However, the nematode dauer larva presents a different situation in which lifespan extension is needed before animals reproduce. Selection for lifespan-extension genes would thus allow the animals to produce progeny. In this case, the *ctl-1* gene may have been acquired for animals to survive extended dauer periods. The adjacency of the *ctl-1* and *ctl-2* genes on chromosome II supports this idea.

Although the cytosolic catalase may be a relatively new acquisition for *C. elegans* and other nematodes, an alternative hypothesis is that cytosolic catalases may be quite general and used particularly during periods of nutrient reduction. Yeast¹⁸ and plants¹⁹ have cytosolic catalases. Although no mammalian gene for a cytosolic catalase has been reported, suggestive evidence exists for cytosolic catalase in mammalian hepatocytes^{20–23}. If confirmed, these observations suggest that cytosolic catalases may be widespread.

The *ctl-1* catalase is elevated during the dauer period, which is induced as a response to overcrowding and depletion of food. The yeast cytosolic catalase and one of the two catalases in *Escherichia coli* are upregulated when these organisms enter stationary phase, a period that is also induced by overcrowding and nutrient depletion^{24,25}. These similarities could indicate that catalases in other metazoans are induced by starvation and similar conditions. Caloric restriction, which extends lifespan in a variety of species, induces catalase activity in rats²⁶. The nature of this catalase activity, whether cytosolic or peroxisomal, is not known. An intriguing speculation is that caloric restriction does not extend life by reducing or altering metabolism, but rather by inducing cell-protective enzymes, like the *ctl-1* catalase, that produce healthier cells. □

Methods

Maintenance of nematode strains. Animals were usually grown at 20 °C (ref. 27). Mutations used were *fer-15*(*b26ts*)II, *age-1*(*hx546*)II, *age-1*(*mg44*)II, *ctl-1*(*u800*)II, *unc-52*(*e444*)II, *daf-2*(*e1370*)III, *clk-1*(*e2519*)III, *daf-12*(*m25*)X and *mec-10*(*u20*). The *age-1* strain TJ401 also contained the *fer-15* mutation; strain DH26 has the *fer-15* mutation alone.

Catalase assays. Catalase assays²⁸ were done on animals killed at various times from cultures synchronized by hypochlorite treatment and washed five

times with M9 buffer²⁷. Lysates were prepared by grinding animals with 0.5-mm glass beads in 20 mM Tris-HCl (pH 7.5), 50 mM potassium acetate, 2 mM EDTA and 100 mM sorbitol.

Molecular biology. General molecular biology methods were as described²⁹. We identified the full-length cDNAs corresponding to the *ctl-1* and *ctl-2* mRNAs by screening a Lambda-Zap cDNA library provided by R. Barstead using the entire expressed-sequence tag cm20b12 (ref. 4) as a probe. A *ctl-1* cDNA plasmid, TU#527, was unidirectionally deleted using the Erase-A-Base system (Promega) and the resulting DNAs were sequenced on an Applied Biosystems model 373. A *ctl-2* cDNA plasmid, TU#528, was sequenced from the 5' and 3' ends and found to be identical to CECAT, a *C. elegans* cDNA encoding a putative peroxisomal catalase that was deposited in the EMBL database by K. J. Henkle-Duehrsen.

We prepared RNA for northern blots from synchronous populations of animals of *daf-2(e1370)* dauers grown at 25 °C. Total RNA was prepared²⁹ from animals washed five times with M9 buffer to remove bacteria. Membranes were probed with *ctl-1*- or *ctl-2*-specific ³²P-labelled riboprobes²⁹ complementary to the 3'-most 264 nucleotides of the *ctl-1* open reading frame or 3'-most 380 nucleotides of the *ctl-2* open reading frame, respectively.

Identification and characterization of *ctl-1(u800)*. During experiments with TU1061 (C.M. and M.C., unpublished data), we identified four clones encoding a cDNA with reduced expression compared to wild type, which were from the *ctl-1* gene. Initial assignment of the *ctl-1* cDNA to the yeast artificial chromosome grid of *C. elegans* genomic DNA indicated that it was located near *ctl-2*. We outcrossed the TU1061 strain based on this map information and obtained a strain exhibiting a significant reduction in lifespan. Genetic mapping of the effect on lifespan placed the mutation in the correct position (between *lin-7* and *unc-52*) on chromosome II.

To identify the site of the *ctl-1(u800)* mutation, we sequenced cDNA amplified by RT-PCR for the entire *ctl-1* open reading frame, which was obtained using the Marathon cDNA amplification kit (Clontech).

Determination of lifespan. Synchronous populations of worms were placed on 60 mm plates seeded with OP50-1 as two-day-old larvae (8–12 individuals per plate, 100–150 individuals per experiment) and grown at 20 °C. Individuals were moved to fresh plates daily for nine days after hatching and weekly thereafter. We examined the plates at various times for live, dead and missing individuals (missing animals were not included in the calculation of lifespan). Individuals were scored as dead when they became unresponsive to prodding with a wire probe and no pharyngeal pumping could be detected¹¹.

Measurement of autofluorescence. Autofluorescent material from *C. elegans* was measured as modified from ref. 6. Synchronous populations of *fer-15*, *fer-15 age-1*, and *fer-15 ctl-1* animals were grown for three days after hatching at 20 °C. The resulting young adults were transferred to 100-mm NGM plates²⁷ containing 40 μM 5-fluoro-2'-deoxyuridine (Sigma) and a lawn of concentrated OP50 bacteria, and maintained at 25 °C. Samples, each collected from three densely populated plates, were washed three times with distilled water and frozen at –80 °C. The fluorescence (excitation at 360 nm; emission at 415 nm) of the 14,000g supernatant from methanol:distilled H₂O (2:1) extracts was measured in a Photon Technology International SFA-20 fluorometer (Hi-Tech).

Construction and analysis of *gfp-ctl* fusions. We generated amino-terminal GFP::CTL fusions for *ctl-1* and *ctl-2* by inserting the appropriate restriction fragments (produced by using PCR) into pPD95.77 (a gift from A. Fire), which contains the gene for the green fluorescent protein with added introns. The fusions were then placed into pPD49.78, which has an *hsp-16-2* heat-shock promoter³⁰, generating the expression plasmids TU#534 (for *ctl-1*) and TU#535 (for *ctl-2*). These plasmids were transformed into wild-type *C. elegans*³⁰ with the dominant *rol-6* plasmid pRF4. Animals were examined after a 1 h recovery following heat shock at 33 °C for 1 h.

Received 25 January; accepted 15 March 1999.

1. Riddle, D. L. & Alberts, P. S. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 739–768 (Cold Spring Harbor Press, New York, 1997).
2. Kenyon, C. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 791–814 (Cold Spring Harbor Press, New York, 1997).
3. Vanfleteren, J. R. Oxidative stress and ageing in *Caenorhabditis elegans*. *Biochem. J.* **292**, 605–608 (1993).
4. Waterston, R. *et al.* A survey of expressed genes in *Caenorhabditis elegans*. *Nature Genet.* **1**, 114–123 (1992).

5. Gould, S. J., Keller, G.-A., Hosken, N., Wilkinson, J. & Subramani, S. A conserved tripeptide sorts proteins to peroxisomes. *J. Cell Biol.* **108**, 1657–1664 (1989).
6. Hosokawa, H. *et al.* Rapid accumulation of fluorescent material with aging in an oxygen-sensitive mutant *mev-1* of *Caenorhabditis elegans*. *Mech. Ageing Dev.* **74**, 161–170 (1994).
7. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. *Science* **277**, 942–946 (1997).
8. Morris, J. Z., Tissenbaum, H. A. & Ruvkun, G. A phosphatidylinositol-3-OH kinase family member regulating longevity and diapause in *Caenorhabditis elegans*. *Nature* **382**, 536–539 (1996).
9. Ogg, S. *et al.* The Forkhead transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* **389**, 994–999 (1997).
10. Lin, K., Dorman, J. B., Rodan, A. & Kenyon, C. *daf-16*: An HNF-3/forkhead family member that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319–1322 (1997).
11. Larsen, P., Albert, P. S. & Riddle, D. L. Genes that regulate both development and longevity in *Caenorhabditis elegans*. *Genetics* **139**, 1567–1583 (1995).
12. Ewbank, J. J. *et al.* Structural and functional conservation of the *Caenorhabditis elegans* timing gene *clk-1*. *Science* **275**, 980–983 (1997).
13. Beckman, K. B. & Ames, B. N. The free radical theory of aging matures. *Physiol. Rev.* **78**, 547–581 (1998).
14. Orr, W. C. & Sohal, R. S. Extension of life-span by overexpression of superoxide dismutase and catalase in *Drosophila melanogaster*. *Science* **263**, 1128–1130 (1994).
15. Amstad, P. *et al.* The balance between Cu, Zn-superoxide dismutase and catalase affects the sensitivity of mouse epidermal cells to oxidative stress. *Biochemistry* **30**, 9305–9313 (1991).
16. Parkes, T. L. *et al.* Extension of *Drosophila* lifespan by overexpression of human *SOD1* in motoneurons. *Nature Genet.* **10**, 171–175 (1998).
17. Kirkwood, T. B. L. & Rose, M. R. Evolution of senescence: late survival sacrificed for reproduction. *Phil. Trans. R. Soc. Lond. B* **332**, 15–24 (1991).
18. Ruis, H. & Koller, F. in *Oxidative Stress and the Molecular Biology of Antioxidant Defenses* (ed. Scandalios, J. G.) 309–342 (Cold Spring Harbor Press, New York, 1997).
19. Scandalios, J. G., Guan, L. & Polidoros, A. N. in *Oxidative Stress and the Molecular Biology of Antioxidant Defenses* (ed. Scandalios, J. G.) 343–406 (Cold Spring Harbor Press, New York, 1997).
20. Roels, F. Cytochemical demonstration of extraperoxisomal catalase. I. Sheep liver. *J. Histochem. Cytochem.* **24**, 713–724 (1976).
21. Roels, F., de Coster, W. & Goldfischer, S. Cytochemical demonstration of extraperoxisomal catalase. II. Liver of rhesus monkey and guinea pig. *J. Histochem. Cytochem.* **25**, 157–160 (1977).
22. Yamamoto, K., Volk, A., Hashimoto, T. & Fahimi, H. D. Catalase in guinea pig hepatocytes is localized in cytoplasm, nuclear matrix and peroxisomes. *Eur. J. Cell Biol.* **46**, 129–135 (1988).
23. Bullitta, C., Ganea, C., Fahimi, H. D. & Volk, A. Cytoplasmic and peroxisomal catalases of the guinea pig liver: evidence for two distinct proteins. *Biochim. Biophys. Acta* **1293**, 55–62 (1996).
24. Bissinger, P. H., Wieser, R., Hamilton, B. & Ruis, H. Control of *Saccharomyces cerevisiae* catalase T gene (CTT1) expression by nutrient supply via the RAS-cyclic AMP pathway. *Mol. Cell Biol.* **9**, 1309–1315 (1989).
25. Loewen, P. in *Oxidative Stress and the Molecular Biology of Antioxidant Defenses* (ed. Scandalios, J. G.) 273–308 (Cold Spring Harbor Press, New York, 1997).
26. Masoro, E. J., Shimokawa, I. & Yu, B. P. Retardation of the aging processes in rats by food restriction. *Ann. NY Acad. Sci.* **621**, 337–352 (1991).
27. Sulston, J. & Hodgkin, J. in *The Nematode Caenorhabditis elegans* (ed. Wood, W. B.) 587–606 (Cold Spring Harbor Press, New York, 1988).
28. Peters, T. J., Muller, M. & DeDuve, C. Lysosomes of the arterial wall. I. Isolation and subcellular fractionation of cells from normal rabbit aorta. *J. Exp. Med.* **136**, 1117–1139 (1972).
29. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
30. Mello, C. & Fire, A. *Methods Cell Biol.* **48**, 451–482 (1995).

Acknowledgements. This work was supported by the NIH (M.C. and J.R.) and by the American Cancer Society (J.R.). Some strains were received from the *Caenorhabditis* Genetics Center, which is funded by the NIH National Center for Research Resources.

Correspondence and requests for materials should be addressed to M.C. (e-mail: mc21@columbia.edu).

Gene silencing in *Neurospora crassa* requires a protein homologous to RNA-dependent RNA polymerase

Carlo Cogoni & Giuseppe Macino

Dipartimento di Biotecnologie Cellulari ed Ematologia, Sezione di Genetica Molecolare, Università di Roma La Sapienza, Viale Regina Elena 324, 00161 Roma, Italy

In plants and fungi, the introduction of transgenes can lead to post-transcriptional gene silencing^{1,2}. This phenomenon, in which expression of the transgene and of endogenous genes containing sequences homologous to the transgene can be blocked, is involved in virus resistance^{3–5} and genome maintenance^{6,7}. Transgene-induced gene silencing has been termed quelling in *Neurospora crassa* and co-suppression in plants. Quelling-defective (*qde*) mutants of *N. crassa*, in which transgene-induced gene silencing is impaired, have been isolated⁸. Here we report the

cloning of *qde-1*, the first cellular component of the gene-silencing mechanism to be isolated, which defines a new gene family conserved among different species including plants, animals and fungi. The *qde-1* gene product is similar to an RNA-dependent RNA polymerase found in the tomato⁹. The identification of *qde-1* strongly supports models that implicate an RNA-dependent RNA polymerase in the post-transcriptional gene-silencing mechanism. The presence of *qde-1* homologues in a variety of species of plants and fungi indicates that a conserved gene-silencing mechanism may exist, which could have evolved to preserve genome integrity and to protect the genome against naturally occurring transposons and viruses.

Post-transcriptional gene silencing (PTGS) following transgene introduction has received a great deal of attention because of its implications in both basic and applied research. A surprising feature of PTGS is that silencing is mediated by a diffusible *trans*-acting molecule in both plants and fungi^{4,10,11}. Several models propose that qualitatively different RNA molecules, defined as aberrant RNAs (abRNAs), which are able to activate gene silencing, may be produced directly from transgene templates or as a consequence of endogenous gene-transgene DNA pairing^{12,13}. The aberration in abRNA could be double strandedness¹⁴, premature termination, or other undefined features. Nevertheless, almost all models assume that short complementary RNA molecules are synthesized by specialized cellular RNA-dependent RNA polymerases (RdRPs) which use the overexpressed or aberrant RNAs as templates^{15,16}. Such cRNA molecules could lead to the formation of double-stranded RNA and eventually to messenger RNA degradation. However, the proposed mechanisms for PTGS are all highly speculative, as no cellular components of the gene-silencing mechanism have been identified previously. To try to identify such components, we began a genetic dissection of the gene-silencing mechanism (quelling) in the fungus *Neurospora crassa*^{17,18}. Using the albino-1 (*al-1*) gene essential for carotenoid biosynthesis¹⁹ as a simple visible reporter for gene silencing, we identified three classes of mutant (*qde*) which were defective in the quelling mechanism⁸.

To clone the *qde* genes, we used random insertional mutagenesis on an *al-1* transgenic strain showing an albino (white) phenotype as a consequence of post-transcriptional silencing of the endogenous *al-1* gene. Of 100,000 independent insertional transformants, one strain (107) showed release of gene silencing, visible as recovery of an orange wild-type phenotype. Using heterokaryon genetic analysis, we assigned the mutation in strain 107 to one of the three previously identified *qde* complementation groups⁸, indicating that strain 107 is mutated in the *qde-1* gene.

To isolate *qde-1*, the tagging plasmid was recovered from the 107 strain by a plasmid-rescue procedure using the restriction enzyme *Bgl*II, which has a single restriction site within the tagging plasmid. One plasmid, pCR4, was recovered after religation of the chromosomal DNA following restriction with *Bgl*II (Fig. 1a). We isolated chromosomal DNA flanking the integration site using the enzymes *Bgl*II and *Pst*I, and the resulting restriction fragment was used as a probe on a *N. crassa* genomic cosmid library. Two positively hybridizing cosmids, 56G11 and 40H7, were introduced by a transformation experiment into strain 107 and into the ultraviolet-induced *qde-1* mutant strain M7. Both cosmids complemented the two *qde-1* mutants tested, restoring *al-1* gene silencing and leading to a white phenotype, indicating that both cosmids contain a functional *qde-1* gene. Using the same flanking DNA as a probe, two *Eco*RI fragments from 40H7 and 56G11 (7.9 and 10 kilobases (kb), respectively) were identified and subcloned (Fig. 1a). Both *Eco*RI fragments (p79E and p10E plasmids) were able to complement two independent *qde-1* mutant strains, but not *qde-2* and *qde-3* mutants, indicating that a functional *qde-1* gene is contained in the 7.9-kb *Eco*RI fragment. The fact that the *qde-1* gene restores *al-1* gene silencing in only the corresponding mutants excludes the possibility that the cloned DNA fragment can reactivate *al-1* gene

silencing irrespective of the *qde* complementation group. The 7.9-kb *Eco*RI fragment was further subcloned by using the *Xba*I site (Fig. 1a), in the middle of the *Eco*RI fragment. Neither of the *Xba*I/*Eco*RI fragments (pSX and pDX plasmids) complemented the *qde-1* mutants, indicating that the *Xba*I site is probably located within the *qde-1* gene.

The region of the *Eco*RI fragment encompassing the putative *qde-1* gene and adjacent regions was sequenced, revealing a long open reading frame (ORF) of 4,206 base pairs (bp) coding for a putative protein of 1,402 amino acids. Two results indicate that this ORF is coincident with the *qde-1* gene. First, the *Xba*I restriction site used to subclone the *Eco*RI 7.9-kb fragment is located in the middle of the ORF (Fig. 1a), consistent with the result that both *Xba*I/*Eco*RI fragments fail to complement the *qde-1* mutation. Second, we mapped the tagging plasmid-insertion site in strain 107 between *Bgl*II and *Nae*I restriction sites within the ORF by Southern analysis (Fig. 1b). No size changes have been detected in the surrounding regions, excluding the possibility that deletions affecting other ORFs within the *Eco*RI 7.9-kb region may have occurred. The *qde-1* mRNA steady-state level was two-fold higher in an *al-1* silenced strain than in a wild-type untransformed strain, indicating that a regulatory mechanism exists that can activate cellular components of the silencing machinery in transgenic strains.

The QDE-1 protein deduced from the nucleotide sequence is composed of 1,402 amino acids and has a relative molecular mass of 158,004: it does not contain a signal peptide or a transmembrane

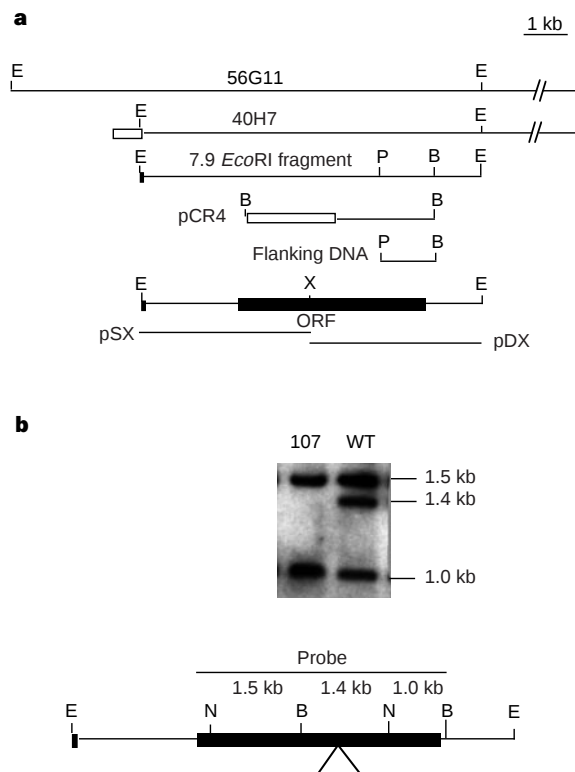


Figure 1 Genomic organization of the *qde-1* gene. **a**, Two cosmids (56G11 and 40H7) that can complement *qde-1* mutants. The white box in 40H7 represents sequences from the cosmid vector. A restriction map of the 7.9-kb *Eco*RI fragment from 40H7, containing the functional *qde-1*, is shown: E, *Eco*RI; P, *Pst*I; B, *Bgl*II. The black box represents the identified ORF within the 7.9-kb *Eco*RI fragment. pDX and pSX plasmids containing *Xba*I (X) and *Eco*RI (E) subcloned DNA fragments are also shown. **b**, Southern blot analysis of wild type (WT) and strain 107. Genomic DNA was double-digested with *Bgl*II and *Nae*I restriction enzymes. The diagram shows the DNA probe used for hybridization and the expected *Bgl*II/*Nae*I (B/N) restriction fragments. The triangle represents the integration site in strain 107 that leads to the disappearance of the 1.0-kb restriction fragment.

domain, indicating that it is likely to be an intracellular protein, and the hydropathy plot indicates that QDE-1 is a soluble protein. A BLAST search revealed that *QDE-1* shares statistically significant homology with hypothetical proteins from several other organisms, including two ORFs of *Caenorhabditis elegans* (accession numbers Z48334 and Z78419), an ORF of *Schizosaccharomyces pombe* (accession number Z98553) and four ORFs of *Arabidopsis thaliana* (accession numbers AF080120 and AC005169 (three ORFs at the same chromosomal location)). Finally, significant homology (expected value 2×10^{-17}) was found with the putative protein encoded by tomato cDNA (accession number Y10403). This homology does not extend over the entire protein, but is restricted to a 570-amino-acid portion defining a conserved domain (Fig. 2). Of the homologous putative proteins identified, only the tomato cDNA has been functionally characterized and shown to encode an RdRP⁹.

Although RdRPs have been considered to be key enzymes in PTGS, no indication of a specific involvement of these enzymes has been provided up to now. Our finding that the *qde-1* gene encodes a putative protein homologous to an RdRP constitutes experimental

evidence for the involvement of an RdRP in PTGS. The presence of homologous genes conserved among different and distant species such as *C. elegans* and *S. pombe* may indicate that gene-silencing mechanisms are present in a wide range of organisms. On the other hand, the presence of several paralogous RdRP genes in the same organism, as have been found in *C. elegans* and *A. thaliana*, may also indicate that these proteins have distinct specialized functions and could be involved in other cellular processes, perhaps using different RNAs as substrates. Aberrant transgenic or viral RNAs that have accumulated after transgenic or viral infection have been proposed to be the substrates of RdRPs specific for gene silencing in plants and fungi^{18,20}. On the other hand, double-stranded RNAs have been postulated²¹ to serve as the RdRP template in the interference phenomenon in *C. elegans*, in which expression of an individual gene can be specifically reduced by microinjecting a corresponding fragment of double-stranded RNA^{22,23}. Identification of the specific RNA substrates for RdRP will help our understanding of the PTGS mechanism and other gene-silencing phenomena. The existence of homologous RdRP genes in different species and the observation that PTGS represents a natural mechanism for protection against

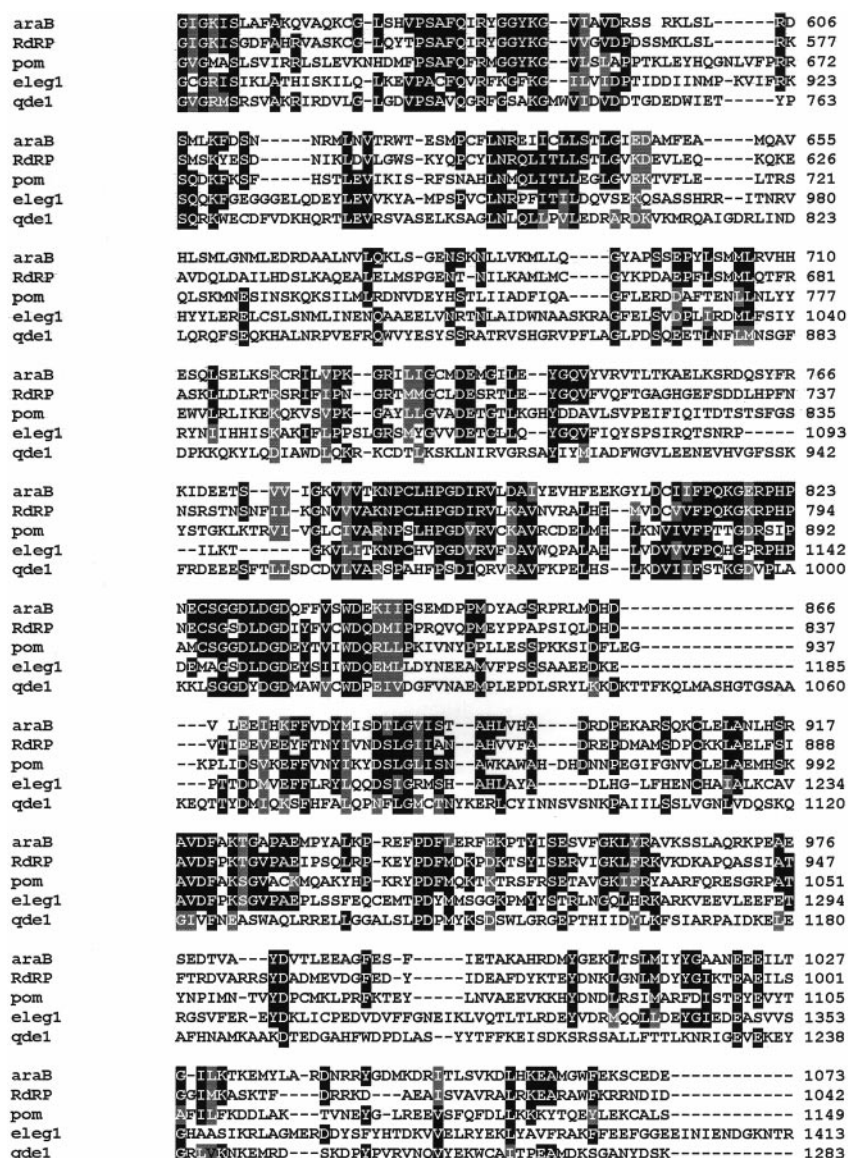


Figure 2 Sequence alignment of the QDE1 protein with other polypeptides from the SwissProt sequence database. ORF of *C. elegans* Z48334 (eleg1), ORF of *S. pombe* Z98553 (pom), ORF of *A. thaliana* AF080120 (araB) and RdRP of tomato

Y10403 (RdRP). Identical residues are shown in black and conservative substitutions are shown in grey.

viruses and invading DNA²⁴ in plants could also indicate that such protection mechanisms may have been conserved during evolution. □

Methods

Strains, growth conditions and transformation. The *N. crassa* methodology and heterokaryon analysis have been described²⁵. Spheroplasts were prepared as described²⁶. Strain 107 was isolated as follows: an *al-1* silenced strain⁸, 6xw, was transformed with pMXY2, which contains a benomyl-resistance β -tubulin gene that functions as a dominant selectable marker in *N. crassa*²⁷. Benomyl-resistant transformants were selected and scored for carotenogenesis by visual inspection of conidial colour: wild-type specimens were orange, whereas white-to-yellow transformants indicated silencing.

Plasmids and libraries. We isolated the genomic *qde-1* gene from a cosmid library of *N. crassa*²⁸. The subcloning of restriction fragments from the genomic library clones was performed in the plasmid pBSK. Subsequently, the subclones were used in co-transformation experiments using pMXY2 or pES200 (bearing hygromycin resistance).

Southern and northern hybridizations. We prepared chromosomal DNA as described²⁹, and after digestion the genomic DNA was blotted as described³⁰. Probes were labelled by random priming (Boehringer). The RNA was electrophoresed on agarose gels, transferred onto Hybond-N membranes and probed.

DNA sequencing and analysis. We determined the nucleotide sequence of both strands of *qde-1* using the *Taq* FS polymerase and fluorescent method, and analysed it using an automatic Applied Biosystems 373A sequencer. We analysed the nucleotide and derived amino-acid sequences using the MacMolly Tetra program. Protein comparisons were made with BLASTP searches. ClustalW algorithm was used for the alignment.

Received 8 January; accepted 16 March 1999.

1. Flavell, R. B. Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc. Natl Acad. Sci. USA* **91**, 3490–3496 (1994).
2. Depicker, A. & Van Montagu, M. Post-transcriptional gene silencing in plants. *Curr. Opin. Cell Biol.* **9**, 373–382 (1997).
3. Ratcliff, F., Harrison, B. D. & Baulcombe, D. C. A similarity between viral defense and gene silencing in plants. *Science* **276**, 1558–1560 (1997).
4. Voinnet, O. & Baulcombe, D. C. Systemic signalling in gene silencing. *Nature* **389**, 553 (1997).
5. Kasschau, K. D. & Carrington, J. C. A counterdefensive strategy of plant viruses: suppression of posttranscriptional silencing. *Cell* **95**, 461–470 (1998).
6. Assaad, F. F., Lee Tucker, K. & Signer, E. R. Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. *Plant Mol. Biol.* **22**, 1067–1085 (1993).
7. Bingham, P. M. Cosuppression comes to the animals. *Cell* **90**, 385–387 (1997).
8. Cogoni, C. & Macino, G. Isolation of quelling-defective (*qde*) mutants impaired in posttranscrip-

- tional transgene-induced gene silencing in *Neurospora crassa*. *Proc. Natl Acad. Sci. USA* **94**, 10233–10238 (1997).
9. Schiebel, W. *et al.* Isolation of an RNA-directed RNA polymerase-specific cDNA clone from tomato. *Plant Cell* **10**, 2087–2101 (1998).
10. Palauqui, J. C., Elmayan, T., Pollien, J. M. & Vaucheret, H. Systemic acquired silencing: transgene-specific post-transcriptional silencing is transmitted by grafting from silenced stocks to non-silenced scions. *EMBO J.* **16**, 4738–4745 (1997).
11. Cogoni, C. *et al.* Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA–DNA interactions or DNA methylation. *EMBO J.* **15**, 3153–3163 (1996).
12. Stam, M., Mol, J. N. M. & Kooter, J. M. The silence of genes in transgenic plants. *Ann. Bot.* **79**, 3–12 (1997).
13. Voinnet, O., Vain, P., Angell, S. & Baulcombe, D. C. Systemic spread of sequence-specific transgene RNA degradation in plants is initiated by localized introduction of ectopic promoterless DNA. *Cell* **95**, 177–187 (1998).
14. Metzlauff, M., O'Dell, M., Cluster, P. D. & Flavell, R. B. RNA-mediated RNA degradation and chalcone synthase A silencing in *Petunia*. *Cell* **88**, 845–854 (1997).
15. Wassenegeger, M. & Pélissier, T. A model for RNA-mediated gene silencing in higher plants. *Plant Mol. Biol.* **37**, 349–362 (1998).
16. Baulcombe, D. C. RNA as a target and an initiator of post-transcriptional gene silencing in transgenic plants. *Plant Mol. Biol.* **32**, 79–88 (1996).
17. Romano, N. & Macino, G. Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.* **6**, 3343–3353 (1992).
18. Cogoni, C. & Macino, G. Conservation of transgene-induced post-transcriptional gene silencing in plants and fungi. *Trends Plant Sci.* **2**, 438–443 (1997).
19. Schmidhauser, T. J., Lauter, F. R., Russo, V. E. A. & Yanofsky, C. Cloning, sequence, and photoregulation of *al-1*, a carotenoid biosynthetic gene of *Neurospora crassa*. *Mol. Cell Biol.* **10**, 5064–5070 (1990).
20. Lindbo, J. A., Silva-Rosales, L., Proebsting, W. M. & Dougherty, W. G. Induction of a highly specific antiviral state in transgenic plants: implications for regulation of gene expression and virus resistance. *Plant Cell* **5**, 1749–1759 (1993).
21. Montgomery, M. K. & Fire, A. Double-stranded RNA as mediator in sequence-specific genetic silencing and co-suppression. *Trends Genet.* **14**, 255–258 (1998).
22. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
23. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
24. Matzke, M. A., Matzke, A. J. M. & Eggleston, W. B. Paramutation and transgene silencing: a common response to invasive DNA? *Trends Plant Sci.* **1**, 382–388 (1996).
25. Davis, R. H. & DeSerres, F. J. Genetic and microbiological research techniques for *Neurospora crassa*. *Methods Enzymol.* **17**, 79–143 (1970).
26. Vollmer, L. J. & Yanofsky, C. Efficient cloning of genes of *Neurospora crassa*. *Proc. Natl Acad. Sci. USA* **83**, 4869–4873 (1986).
27. Staben, C. *et al.* Use of a bacterial Hygromycin B resistance gene as a dominant marker in *Neurospora crassa* transformation. *Fungal Genet. Newslett.* **36**, 79–81 (1989).
28. Cabibbo, A., Sporeno, E., Macino, G. & Ballario, P. CBM1, an *Neurospora crassa* genomic library in pAC3 and its use for walking on chromosome VII right arm. *Fungal Genet. Newslett.* **38**, 68–70 (1991).
29. Morelli, G., Nelson, M. A., Ballario, P. & Macino, G. Photoregulated carotenoid biosynthetic genes of *Neurospora crassa*. *Methods Enzymol.* **214**, 412–424 (1993).
30. Maniatis, S. T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1982).

Acknowledgements. We thank G. Azzalin for technical assistance, and G. Coruzzi and A. Pickford for comments on the manuscript. This work was supported in part by grants from the Istituto Pasteur Fondazione Cenci Bolognietti, from the European Union BIOTECH program and from the Ministero dell'Università e della Ricerca Scientifica e Tecnologica.

Correspondence and requests for materials should be addressed to G.M. (e-mail: Macino@bce.med.uniroma1.it). The *qde-1* gene sequence has been deposited with Genebank (accession number AJ133528).

Cell cycles to cell markers

Smaller (volumes for fluorescence spectroscopy), faster (mRNA assays) and bigger (inventories of antibodies) – more new and refurbished products for molecular and other biologists vie for your attention.

Cell-cycle reagents

From BioSource International

www.biosource.com

Direct conjugates for flow cytometry

Additions to the CycleVision range include FITC conjugates for cyclin D1, cyclin B1, p53, p21^{WAF1} and p27^{Kip1}. These directly conjugated monoclonals are designed to eliminate the need for a secondary antibody for flow cytometry applications, simplifying the procedure. All products are monoclonal antibodies, produced in mouse, conjugated to FITC. Data sheets feature suggested protocols for flow cytometry, including working dilutions, positive controls and references. The antibodies have been tested and characterized for use in applications including western blotting, flow cytometry, immunoprecipitation and immunohistochemistry.

Reader Service No. 100

Buffers

From Rockland Immunochemicals

www.rockland-inc.com

A buffers' guide to molecular biology

Rockland Immunochemicals Inc., producer of over 3,000 antibodies, has now introduced a range of buffers listed in the leaflet "Buffers for molecular biology". All solutions are manufactured under cGMP guidelines, and are sold as DNase- and RNase-free and sterile. The buffers are prepared to be free of endogenous protease and nuclease activity and endotoxins. Buffers come in 100-ml and 1-litre sizes and the range includes DEPC-treated water, Tris-HCl, TE, PBS, TBS, EDTA and sodium acetate.

Reader Service No. 101

Piceatannol

From Kamiya

www.kamiyabiomedical.com

Syk protein inhibitor for signal transduction research

Syk kinase is a cytoplasmic, 72 kDa non-receptor kinase that becomes tyrosine-phosphorylated prior to activation. It mediates signalling by the T-cell receptor, B-cell receptor, IgE receptors and G proteins, and has been associated with mast cell activation, platelet activation and cell adhesion. Piceatannol (3,4,3'-5'-tetrahydroxyl-trans-stilbene) is a membrane-penetrating inhibitor of the Syk protein-tyrosine kinase. It also inhibits a 40-kDa protein-tyrosine kinase isolated from thymocytes, but has little effect on serine/threonine kinases.

Reader Service No. 102

MultiPROBE II

From Packard Instruments

www.packardinst.com

New accessories for molecular biology applications

The MultiPROBE II EXpanded Robotic liquid-handling system has two new modules. The first is a vacuum manifold control system to automate DNA extraction and plasmid preparation using commercially available kits. All liquid-handling steps and control of the vacuum pump are performed by the WinPREP software. The second module is an integration package for thermal cyclers. Full DNA extraction and amplification assays can be performed including extraction, assay set-up, sample/reagent transfer and cycle sequencing.

Reader Service No. 103



Roche cocktail: keep taking the protease inhibitor.

Complete protease inhibitor

From Roche

www.roche.com

Protection for proteins, in tablet form

When added to aqueous solutions — including cell lysis buffers used during the manipulation of proteins — the water-soluble Complete Protease Inhibitor Cocktail Tablet releases stable and consistent doses of inhibition. These tablets contain an optimized mixture of non-toxic protease inhibitors that inhibit proteases including serine, cysteine and metalloproteases. The tablets are available in two sizes, for 10-ml and 50-ml solutions, and with or without EDTA.

Reader Service No. 104

AR441

From Dako Corporation

www.dakousa.com

Monoclonal anti-androgen receptor, AR441

Androgen receptor AR441 is an intracellular protein that mediates the actions of androgen steroids, testosterone and 5- α -dihydro-

ADVERTISEMENTS

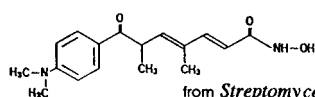
**Customised
monoclonal &
polyclonal
antibodies
with
guarantee!**

BioGenes GmbH

Köpenicker Str. 325 · D-12555 Berlin
Tel. +49-30-65 76 23 96 Fax: - 97
E-mail: BioGenes@Berlin.sireco.net
Internet: <http://www.biogenes.de>

READER ENQUIRY NO. 17

Trichostatin A



from *Streptomyces*

Catalog No. 204-11991

This histone deacetylase inhibitor is highly effective at low concentrations, with high specificity. For more information, please call

Wako BioProducts: (800) 992-9256; (804) 271-7677.

READER ENQUIRY NO. 98

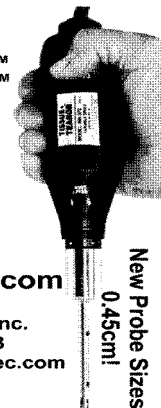
Homogenizers?

- Tissue-Tearor™
- BeadBeater™
- Mini-BeadBeater™
- BioHomogenizer™

**Check
Us
Out**

www.biospec.com

BioSpec Products, Inc.
Ph/Fax 918-336-3363
e-mail info@biospec.com



New Probe Sizes
0.45cm!

READER ENQUIRY NO. 54



imMedia™ – Fast Food for *E. coli*

imMedia™ is pre-mixed, presterilized *E. coli* growth medium that can save you hours. With imMedia™ you can prepare medium immediately because there's:

- No mixing of media components
- No autoclaving
- No adding antibiotics, IPTG, or X-gal

Simply add water and microwave for 3.5 minutes and you are ready to go. When you need media in a hurry, think imMedia™ from Invitrogen. Fast Food for *E. coli*.



Available in eight different flavors.

www.invitrogen.com
1600 Faraday Avenue • Carlsbad, CA 92008
1-800-955-6288



READER ENQUIRY NO. 12

testosterone. Androgen receptor has been shown to be present in the nuclei of most neoplastic cells, with neuroendocrine differentiation in prostate adenocarcinomas. The antibody gives localized nuclear staining and can be used on formalin-fixed, paraffin-embedded tissue sections.

Reader Service No. 105

Cyclic AMP/GMP ELISA

From Biomedical Technologies

www.btiinc.com

Two new ELISA kits

These new kits are for the analysis of cyclic AMP and cyclic GMP in the range 0.05–100 picomoles per millilitre. Levels of cyclic AMP or cyclic GMP can be measured in most biological samples, including urine, plasma and cell culture, without the need for extensive purification. Both assays can be run either overnight or in 6 hours.

Reader Service No. 106

Quantikine mRNA

From R&D Systems

www.rndsystems.com

A new kit for mRNA quantification

R&D's new calorimetric microplate assay quantifies cytokine-specific mRNA at low levels, using total or poly(A)⁺ RNA. The Quantikine mRNA assay takes less than a day and handles high-throughput and multiple-mRNA targets on one microplate. Kits are



Open the box: R&D's mRNA assay.

available for human IL-1β, IL-2 and IL-6.

Reader Service No. 107

G-148

From Promega UK www.euro.promega.com/uk

An aminoglycoside to block protein synthesis

Antibiotic G-418 sulphate blocks protein synthesis in eukaryotes by targeting the 80S ribosome. It's used in molecular genetics as the selective agent for the neomycin resistance gene. The *neoR* gene product, aminoglycoside 3'-phosphotransferase, inactivates G-418 by phosphorylation. G-418 is used to select and maintain stable eukaryotic cell lines that have been transfected with vectors containing *neoR*.

Reader Service No. 108

ThioFusion

From Invitrogen

www.invitrogen.com

An expression vector for high protein yields

The pBAD/TOPO ThioFusion Expression

ADVERTISEMENTS

NEWCASTLE UNIVERSITY
MOLECULAR BIOLOGY UNIT



PROTEIN SEQUENCING

Membrane-bound, solid or solution samples

£12 per cycle

5 cycles minimum - no set up charge

For more information please contact Dr. Joe Gray:

Phone/Fax: +44 191 222 8612

Email: JOE.GRAY@newcastle.ac.uk

DNA sequencing (ABI 377-96KL, 877 Catalyst)
also available at competitive prices

READER ENQUIRY NO. 2

The Mouse Kit™ for Gene Targeting in ES Cells

- ✓ Complete with embryonic stem (ES) cells, feeder cells and essential reagents (including FBS) for engineering one knockout mouse strain
- ✓ Low-passage AB2.2-Prime ES Cells® (129/SvEv)
- ✓ Uniform, high-quality ESQ feeder cells (mitotically inactivated STO cells) engineered for *neo* and *puro* resistance
- ✓ Prequalification of ESQ feeder cells and reagents for use with AB2.2-Prime ES Cells

*Manufactured by Lexicon Genetics Incorporated

Fax: (619) 535-0071

Tel: (619) 535-5400 or (800) 424-5444

www.stratagene.com

READER ENQUIRY NO. 6

STERNBERGER MONOCLONAL (SMI) ANTIBODIES FROM AFFINITI

The SMI series of antibodies to neurofilaments, tau, myelin basic protein, blood brain barrier (EBA), SNAP-25, and tubulins is available from AFFINITI.

We also offer many antibodies to proteins of neuroscientific interest, including GAD65/67, α, β and γ (NEW) Synucleins, heme oxygenase, neuronal nitric oxide synthase, PGP9.5 and β-catenin.

For data sheets or more information on AFFINITI's extensive range of products for neuroscience research please visit our website at:

[HTTP://WWW.AFFINITI-RES.COM](http://WWW.AFFINITI-RES.COM)

AFFINITI Research Products Ltd., Mamhead Castle, EXETER, EX6 8HD, UK
Tel. +44/0 1626 891010; Fax: +44/0 1626 891090; URL: <http://www.affiniti-res.com>

READER ENQUIRY NO. 15

System is specifically designed to increase recombinant protein yield in *E. coli* via tightly regulated, soluble expression of recombinant proteins. The kit uses the pBAD/Thio-TOPO expression vector which encodes an N-terminal, cleavable thio-redoxin fusion to improve the solubility of the protein expressed. This vector carries the araBAD promoter for regulated expression with high inducibility and low background. The pBAD/ThioTOPO vector is topoisomerase-I-activated to allow 5-minute TOPO cloning of PCR products.

Reader Service No. 109

LabelIT

From Mirus Corporation

www.genetransfer.com

One-step nucleic-acid labelling

LabelIT non-radioactive labelling reagents are marketed as simple and highly reproducible non-radioactive DNA-labelling tools for a wide variety of molecular and cell biology applications. Cy3 and Cy5 are popular dyes that have broad application in DNA-sequencing detection and other nucleic-acid studies. The LabelIT protocol is simple and avoids the need for enzymatic labelling reactions. The product is labelled DNA and RNA that is intact and ready to use (after a rapid, simple clean-up step) in less than one hour. Incorporation of the label is simply controlled by varying the nucleic-acid or reagent concentrations in the reaction.

Reader Service No. 110

ConfoCor

From Carl Zeiss

www.zeiss.com

A new instrument for measuring biomolecular interactions in microlitre samples

ConfoCor applies fluorescence correlation spectroscopy to the measurement of molecular interactions. Advantages of this technique include the ability to work in very small volumes (down to nanolitres), a broad dynamic range (molar to picomolar), rapid measurement time (seconds), extreme sensitivity (single-molecule detection), insensitivity to background noise, and a facility for self-calibration. Applications of fluorescence



Zeiss work: small volumes, and single molecules.

correlation spectroscopy include ligand-receptor interactions, DNA-protein interactions, DNA-DNA interactions, viral prion and bacterial detection, particle size determination, antigen-antibody interactions and enzyme activity determination.

Reader Service No. 111

Proteasome inhibitor

From Affiniti Research Products

www.proteasome.com

A peptidyl boronic-acid inhibitor for proteasome system research

The leupeptin analogue MG-132 (Z-Leu-Leu-Leu-H) is a potent cell-permeable inhibitor of proteasomal chymotrypsin-like activity. It also induces apoptosis in Molt-4 and L5178Y cells via a p53-dependent pathway. The boronic-acid derivative is a potent proteasome inhibitor with a K_i of 0.023 nM versus SDS-activated 20S proteasome.

Reader Service No. 112

Anti-cytokeratin clone AE3

From BioGenex

www.biogenex.com

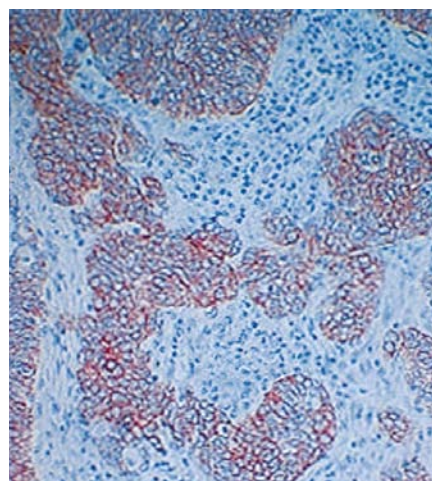
High-molecular-weight cytokeratin monoclonal antibody stains for paraffin tissue sections

Anti-cytokeratin clone AE3 can be used as a broad-spectrum marker for the basic cytokeratins and is particularly suited for distinguishing poorly differentiated carcinomas from other non-epithelial neoplasms. Monoclonal antibody AE3 reacts specifically with all basic, type-II (high molecular weight) cytokeratins in formalin-fixed, paraffin-embedded tissue sections. AE3 antibody is available in concentrated form or as a ready-to-use reagent optimized for use with Super Sensitive detection systems.

Reader Service No. 113

These notes are compiled in the Nature office from information provided by the manufacturers.

For more details, fill in the reader service card bound inside the journal.



Cytokeratin in squamous carcinoma cells.

The New Generation of Online Sequencing

M@XControl™

On-line Sequence Information System

INTERNET BEACH CAFE

MacControl Online Sequence Information System

- ✓ Job No. 38 finished
- ✓ Job No. 39 QC passed
- ✓ Job No. 36 finished

Enjoy your holiday!

It's nice to see you so relaxed!

www.gatc.de

GATC

The WorldWideLab™

GATC GmbH
Fritz-Arnold-Str. 23
D - 78467 Konstanz

Tel. (+49) 75 31 - 81 60 0
Fax (+49) 75 31 - 81 60 81
maccontrol@gatc.de

READER ENQUIRY NO. 511

Out of the lab and into the marketplace

Any researcher who leaves the academic world to start their own company will face a steep learning curve when it comes to the financial and business aspects. *Nature's* writers offer practical advice in this eight-page feature.

Brendan Horton

Start-up companies play an important role in the progression from scientific discovery to the point where society can benefit from its investment in basic research. Small start-up companies are in an environment that allows them to do much of the risky, early-stage development of innovations that come out of academic labs, and to push them through to the large-scale manufacturing arena, in a more efficient manner than is possible for large biotechnology or pharmaceutical companies. Seeing how life-science start-ups position themselves to do business helps to illustrate this process.

For start-ups to be successful, they must develop and maintain academic collaborations that feed a product pipeline with potentially useful technology. They must partner large pharmaceutical companies to which their products may be sold and, just as important, they must satisfy a variety of financial backers in the process. Balancing these three obligations is no small challenge.

In his pursuit of therapeutic treatments for cancer, Nicolas La Thangue has added a level of difficulty to the task, by doing double duty as the chief scientific officer at Prolifix, a cancer and cell-cycle drug discovery company that he co-founded, and as a professor at the University of Glasgow, where he leads a group of cell-cycle researchers. Despite the high demands of both these jobs, he says he is able to maintain a good balance between the two: "The two complement each other very well."

Many scientists claim in their grant applications that they will cure cancer or heart disease, without actually understanding how you do it, says La Thangue. How you move academic results into a product-development line with a view to making a drug is largely unknown to academics, he says. For La Thangue, the chance to explore this process came while he was at the MRC National Laboratory for Medical Research in Mill Hill, London. After making interesting observations about the cell cycle and identifying potential drug targets for tumour cells, he was confronted with the problem of how these results could be exploited therapeutically. "The only way to do that is through the formation of a company, or being involved with a company," he says. However, if you interact with a large company, you have little control over what gets done, whereas in a small company you have a lot of control, he says, particularly if you are a founder.

The essential difference between academic research and biotechnology, says La



No bottling out: Valerio says it's a thrill to be in a company where success brings the chance to grow.

Thangue, is that biotech is looking to bring in new ideas and new targets and move them forward into drug discovery as fast as possible. This work is milestone- and objective-oriented — projects are driven by events that need to happen at a certain time. If the timelines are not met, the company must answer to its investors or partners.

Communication challenge

For La Thangue, the learning curve has been steepest for the financial and business considerations: "I've learned more in the past four years than I had in the whole of my life, having to keep investors happy, to drive projects forward and work to budgets. It's a very different lifestyle." Even though the work may be scientifically excellent and going forward fast, if it is not communicated properly, he says, it may be unappreciated. "Keeping investors in tune with what's going on in the company... as well as keeping teams in the company happy and enthusiastic, requires just about everything that you've learned." This means communicating effectively with a non-technical audience, says La Thangue.

Conversely, those working within the company must understand the science, according to Dinko Valerio, founder and chief executive officer (CEO) of IntroGene, a gene therapy company based on the premises of Leiden University in Leiden's Bioscience Park in The Netherlands. "We have to convince the pharmaceutical companies that our products are great products," which means winning over fellow scientists.

Another factor that is important to

investors and business partners alike is filling out a management team. Currently, Valerio is doing the work of chief scientific officer (CSO) and CEO. But, in the last step to building a management team, IntroGene is searching for a CSO. "It is clear that scientific leadership should be given to somebody who can dedicate all their time to it," says Valerio. Also important is to find the right financial officer and director of business and marketing.

"The enjoyable aspect in a company such as ours is that we have identified a dot on the horizon, and the whole company is working toward that goal," says Valerio. "In our case, that is to develop a meaningful gene therapy product to treat a major human disorder. This dot may be very far away, but it must still be concrete enough for everyone to realize all the steps that are taken are towards that end."

"It's a big thrill to be in an entity where success immediately reflects itself in opportunities to further grow the organization," says Valerio. In the past six months, IntroGene has more than doubled its number of employees, to nearly 100. But, he adds, "recruiting those people and making sure that everyone realizes what the dot on the horizon is... that is a big challenge."

La Thangue says those considering a career outside the academic world should remember that "there's no real chance to do exploratory research, at least in the small companies in the UK". Another consideration, he says, is whether to join a big company, where there is administration and the structure is such that things move pretty slowly, or to become part of a small company

First write your business plan

After completing a degree in chemical engineering and working for six years in industry, David Perry went to Harvard Business School with the goal of being involved with or launching his own high-technology start-up company. "If you can get people who understand technology and people who understand business and put them together so that they communicate with each other effectively, you've got something extraordinarily powerful," says Perry.

He has taken the business plan that earned him runner-up honours in the inaugural contest for Harvard Business School and put it to the test — starting his own company. In 1997 he co-founded Chemdex.com, where he is now the chief executive officer. The company uses e-commerce technology on the Internet to improve the process of finding and buying life-science products for researchers.

According to Perry, communication is often the most difficult part about launching a start-up. Often technologists cannot explain what they are doing or the business people cannot grasp the technology. Although Perry hates to pretend that he is an expert, he says: "I did it once and it worked; I'm not sure how much was luck and how

much was skill." He does stand by some principles that he used in developing Chemdex.com. One is to get as many people's opinions as you possibly can. "Sometimes you learn something, sometimes you don't, but it allows you to refine your business plan as you go. People think of things you haven't thought of or will ask you questions you have to figure out the answers to. In the end, you can either convince people you have a good plan or you can't," says Perry.

He adds that, in organizing a company, you want to hire people who are better than you in every category. "If I don't hire someone who's better than me as a chief financial officer or vice-president of sales or marketing, then I'm not doing my job." The only way to do this, however, is to have enough working knowledge of those areas to tell good from bad, he says. Make sure that you use your skills. If you're really good with technology, make sure you find people who are good at the other things. The fault he sees most often is that technical people go into business and hire other technical people but undervalue financial or management skills. "It's very hard to grow a business unless you have other people who can do those things well," says Perry.

B.H.

where the environment is more dynamic and you are more actively involved in the overall process. "We are a drug discovery company, very focused on assay building and meeting objectives in drug discovery pipelines. When we interview people we make sure they are absolutely aware of it," says La Thangue.

According to Valerio, the work differs from academic research in the absolute focus that you should keep in industry. For the most part, you should leave alone the many paths you will come across, because it's the end goal that you want to reach. On the other hand, flexibility is also important, because "you should be able to move away from something that looks like a dead end," he says.

Generally, projects in start-up companies are driven by budgets, which are forecast over one- to five-year periods. The availability of financing is far more volatile than it is in academic research. Market forces drive the financing and investor confidence, says La Thangue. Whether or not investors see a product or the possibility of a public offering of shares also influences the level of financing enormously. In academic research, support depends essentially on the research and publication record that you have, he says. "The funding is tight, but it's consistently tight." In a biotech company, funding goes up and down, and this is essentially out of the control of the people in the company, he adds.

For example, biotech has gone out of favour in the United States, and it is much

more difficult to get a biotech company started there these days, according to Rolf Schneider-Günter, a venture capitalist who specializes in the life sciences for Atlas Venture in Frankfurt, Germany. Most prominent venture capitalists have shied away from biotech to a large extent or stopped investing in it altogether. Since about 1996, however, the climate in Europe has become much more favourable, apart from in the United Kingdom, where it is the same as in the United States, says Schneider-Günter.

According to *The New York Times* last November, the climate for start-ups has become a "far cry from the feeding frenzy of 1997 and the first half of 1998... start-ups face a market where they have to pitch their business plans to more venture capitalists, and give up more equity in order to raise cash".

No shortage of advice

According to Valerio, the atmosphere today is different from when he launched IntroGene because more and more life-science companies are spinning out of academic research in Europe. "The whole infrastructure of professional business developers that you can use as advisers, the public relations people, the venture capital community, have moved into the life sciences in a much bigger way than in the early days." He says that relationships with academic institutions also needed to be developed. IntroGene was treading on new ground, structuring rela-

tionships that included equity positions for universities in spin-out companies, which is still the model being used in The Netherlands today. There were few role models in those days. "But I never sensed any negative sentiments about 'going commercial'."

US companies are increasingly looking to Europe for available capital. For example, Riboscience, a company based in Boulder, Colorado, has opened a subsidiary in Germany, and the San Francisco-based Exelixis co-founded Artemis Pharmaceuticals in Germany with Nobel laureate Christiane Nüsslein-Volhard. "An increasing number of American companies are talking to us about the possibility of establishing something over here, and ease of financing is the main reason," says Atlas's Schneider-Günter. The climate has improved considerably thanks to the establishment of technology-oriented stock exchanges, such as EASDAQ, Le Nouveau Marché and Neuer Markt, says Schneider-Günter. By some accounts Europeans in general are bullish about start-up companies, if not down right speculative. (For more information on venture capital markets, search *The New York Times* website with the key words 'venture capital'.)

In Germany in particular, says Schneider-Günter, the combination of venture-capital mechanisms, the technology stock exchanges, a government push into biotechnology, substantial funding, and the availability of 'soft' money, has provided a framework that is very conducive for early stock offerings. □

Brendan Horton is a freelance journalist.
e-mail: b.horton@naturedc.com

Wisdom on the web

- **Convergence Partners**
<http://www.convergencepartners.com/>
- **Walden Group**
<http://www.waldenvc.com/>
- **New York Times search engine**
<http://search.nytimes.com/search/daily/>
- **APS on physics start-ups**
<http://positron.aps.org/meet/CENT99/BAPS/tocF.html#SFC08.001>
- **Oxford Science Park**
<http://www.oxfordsp.com/>
- **Atlas Venture Capital**
<http://www.atlasventure.com/>
- **IntroGene**
<http://www.introgene.com/>
- **Prolifix**
www.prolifix.co.uk
- **Chemdex**
<http://www.chemdex.com>
- **The Whitaker Foundation**
<http://www.whitaker.org/>
- **European Life Sciences Conference**
<http://www.elsc.com/>
- **MRC Technology Transfer Group**
<http://www.mrc.ac.uk/ttg.html>

'Venture fever' grips Asian economies

David Swinbanks

The economic crisis in Asia has awakened governments to the need to cultivate venture businesses to help pull their countries out of recession and compete with the West. Job opportunities in start-up ventures are likely to expand over the years ahead, first in the more nimble and dynamic economies of Taiwan and Singapore, but later in China, Korea and Japan as well.

Asian scientists trained in the West will be central in the development of these start-ups. A dynamic interaction is developing between Asia and the West, in particular with Silicon Valley in the United States, in which entrepreneurial Asian scientists are using venture capital from both local and Western sources to establish start-ups in Asia and the United States.

Last month, the Singapore government announced the creation of a US\$1 billion fund to stimulate what it calls "technopreneurship" (*Nature* 398, 553; 1999). "Singapore is really getting to grips with the key issues surrounding the growth of high-technology businesses," says Hugh Purser of Cambridge Corporate Consultants in Britain, who formerly worked as an adviser to the National Science and Technology Board in Singapore.

"There is a good understanding of what needs to be done, and a commitment of resources from the very top," says Purser. "Some will say that, for entrepreneurial activity to flourish, the best thing is for govern-



Triangular collaboration: Chen (left) and colleagues at Tsinghua University in China have formed a pharmaceuticals business with US scientific input and finance from Hong Kong (see box on page 178).

ments to get out of the way. But in Singapore's case it is essential for government to take the lead, until a critical mass is achieved."

Japan is also beginning to realize that it has been left behind in the development of new industries, particularly in biotechnology, genomics, bioinformatics and computing. Japan's Ministry of International Trade and Industry plans to invest ¥50 billion (\$400 million) in a supplementary budget this year to support venture businesses as a first step in an ambitious plan to create 1,000 biotechnology-related companies and 70,000–80,000 jobs within the next decade

(*Nature Biotechnology* 17, 320–321; 1999). But the academic community in Japan is generally averse to getting involved in business, and the country will probably continue to lag in this area for many years.

South Korea has been gripped with 'venture fever' since the country plunged into recession in late 1997. The government has introduced laws to allow university professors to set up and head venture companies and use university facilities to make products, and there are already a few success stories. But, as in Japan, many cultural and regulatory hurdles remain.

Singapore aims to become a hot spot for life sciences

Singapore is primed for the launch of start-up companies in the life sciences. Last month, the deputy prime minister Tony Tan Keng Yam announced a US\$1 billion "technopreneurship" fund, signalling the island state's commitment to cultivating venture businesses and venture funds.

This follows other recent initiatives, such as the \$20 million Life Science Investment fund established last November by the National Science and Technology Board and the Economic Development Board to commercialize home-grown and overseas technology in the life sciences. This was followed in December by the new Centre for Drug Evaluation to evaluate new drugs according to international standards.

Singapore is targeting the development of drugs, medical and food products and agrobiotechnology. It has been building strengths in these areas over the past decade. For example, the Institute of Molecular and Cell Biology, established in the late 1980s, has

won wide respect. In 1993, it entered into a joint venture with the British pharmaceutical giant GlaxoWellcome to use molecular and cell technology to screen natural products for drugs. And in 1995 the institute spun-off Singapore's first life-science venture, GeneSing, which is developing health care products for the Asia-Pacific market.

More recently, the Institute of Molecular Agrobiotechnology, established in 1995, has set up a joint venture with Monsanto of the United States to commercialize the use in China of cotton that has been genetically modified to fight bollworm pest (see *Nature* 383, 14; 1996). And a new BioInformatics Center at Singapore National University has created a successful spin-off holding company, Bioinformatrix, and a US-based venture, KRIS Technology, which sells bioinformatics software.

Western investors are sitting up and taking note. One group that has invested in the early stages of life-science ventures in the

United States and Europe has been investigating opportunities in the Far East for the past three years. It has concluded that Singapore is where it is "most comfortable", and is soon due to announce a significant investment. "In addition to the frequently cited attributes of Singapore (transparent legal and accounting practices, English language and education), we are particularly impressed with Singapore's commitment to supporting techno-entrepreneurship, and the city state's outstanding Institute of Molecular and Cell Biology," says one investor of this group.

The Singapore government is well aware that it has to develop a whole spectrum of talent, including not only researchers and entrepreneurs, but also investment bankers, analysts, venture capitalists, and patent and corporate lawyers. It is adopting an open-door policy to draw in talent from around the world, and jobs in these areas will abound in the years ahead.

D.S.



Cheng: forced to look outside China's borders for financial backing for his biochip venture business.

Taiwan is a pioneer in the region in the development of venture businesses. It has made a great success of its Hsinchu science park south of Taipei, which over the past decade and a half has drawn thousands of Taiwanese scientists and engineers back from the West. Dozens of successful start-ups in information technology have been established in the park, many of them linked to Silicon Valley in the United States, where more than one-third of the engineers and researchers are said to be of Taiwanese origin. A push into biotechnology that began at Hsinchu several years ago is also beginning to show some signs of success, which should be helped by substantial new government investments in research on agrobiotechnology, medical genomics and biopharmaceuticals (see box opposite).

Since moves to develop a market economy in China began in the 1980s, hundreds of start-up companies have been established by

institutes of the Chinese Academy of Sciences and universities (*Nature* 378, 540–542; 1995). Most have been failures, but a significant few have been successful and have pumped much needed funds into their mother institutions.

One example is Beida Weiming Biotech Corporation, set up by Zhang-liang Chen, vice-president of Peking University, when he was a department chairman in 1992. The corporation's joint venture company, Kerxing Biopharmaceutical, established with a US company in 1994, now controls 60% of the market for interferon in China, and the corporation will shortly be listed on the Shanghai stock exchange, according to Chen.

But Chen sees many barriers to the successful establishment of start-up companies in China. There is a lack of creativity and originality in China's nascent biotechnology industry, and a tendency just to copy the

West's products. There is a need for greater protection of intellectual property rights, and venture capital is seriously lacking, particularly the long-term type required for biotech start-ups, says Chen.

Government regulations need to be eased, for example, to allow companies to donate money to universities and academic institutions free of tax. And the attitude of the academic community in China, which tends to look down on business activities, needs to change. Last but not least, Chen points to the dearth in China of scientists with entrepreneurial and Western business management skills.

But there is a growing realization in Chinese government circles and the research community that radical reforms are required for venture businesses to thrive. Tsinghua University in Beijing, one of the country's leading universities, at one time had 140 start-up companies, but this has now been slashed to about a dozen successful ones.

Similarly, the 123 institutes of the Chinese Academy of Sciences have more than 500 companies but, according to Yi Xun Yan, one of the four vice-presidents of the academy, only about a third are successful (many would say much less). The rest are a "big problem" for institute directors, says Yan, because it is difficult to close such companies, owing to social obligations to look after the employees.

Nevertheless, the academy has learnt "many important lessons" over the past decade, says Yan, and he is optimistic. He says the government is studying the need to create venture funds, and thinks there is growing awareness of the need to protect intellectual property rights.

But Yan, who established a successful thin-film optics joint venture linked to the acade-

China's scientists tap offshore funds

Two entrepreneurial scientists at Tsinghua University in Beijing who have recently returned to China from the West provide examples of how venture businesses can be established in Asia even when the climate is far from ideal.

George Guo-Qiang Chen carried out microbiological research in Austria, Britain and Canada before returning to China in 1994. In collaboration with Daniel Shao, a financier in Hong Kong, and Thomas Wagner and other scientists at Ohio University in the United States, he has established Huagen Pharmaceutical, which plans to manufacture biopharmaceuticals based on recombinant peptides produced by transgenic *Escherichia coli*. Chen could not find investors in China because of legal hurdles and a "huge gap" in culture with the West, which is why he and his collaborators at Ohio turned to Shao.

Under a triangular arrangement, three people in Hong Kong will raise finance for the venture, four researchers at Ohio are developing recombinant organisms, an area pioneered by Wagner, while researchers at Tsinghua with experience in engineering will set up pilot production facilities. One Tsinghua scientist goes to Ohio each year to interact with the US collaborators.

Similarly, Jing Cheng (above), who has just returned to China to head a new biochip research and development centre at the university, is setting up a biochip venture business. Finance is expected to come from the United States, Taiwan, Hong Kong, and possibly mainland China itself.

Cheng started his career as an electrical engineer in a Chinese locomotive factory in 1983. He then went to Strathclyde University in Scotland to do his PhD and developed and

patented a rapid DNA extractor in the department of pure and applied chemistry, before going to Aberdeen University's department of molecular and cell biology to develop capillary electrophoresis-based methods for mutation detection. He moved to the University of Pennsylvania School of Medicine to develop biochips under the sponsorship of PE Applied Biosystems, and then went to the US company Nanogen where he developed a portable bioelectronic chip for field use (see *Nature Biotechnology* 16, 541–546; 1998).

Cheng was attracted to Tsinghua by the opportunity to do interdisciplinary research spanning science and engineering, and he hopes Tsinghua will become the first university in China to enter high-tech ventures overseas and actively "borrow" foreign experience.

D.S.

my's Shanghai Institute of Technical Physics, is well aware of the difficulties of bringing ideas from the lab to the market. Like Chen, he sees a great need for Chinese scientists trained in Western management skills to return to help China to develop start-ups.

Tsinghua University has recently recruited Jing Cheng from Nanogen, a biochip venture in the United States, to head its new Biochip R&D Centre. Cheng's group will receive over US\$1 million research and development funds from the university, and Cheng is also setting up a biochip start-up company using offshore capital (see box opposite). This approach of seeking venture capital offshore has been adopted by entrepreneurial scientists elsewhere in the region, and it may become a widespread phenomenon as governments struggle to change the regulatory, financial and cultural environment that inhibits venture business.

For example, Sunyoung Kim at the Institute for Molecular Biology and Genetics of Seoul National University, who in late 1996 established ViroMed, South Korea's first venture business in gene therapy and diagnostics, has sought investment and strategic alliances in the United States and Europe because of the lack of long-term venture capital in Korea. And bioinformatics researchers at Singapore National University have set up a company called KRIS Technology in

Menlo, California, to market bioinformatic software products.

It will take time for government initiatives to stimulate venture businesses in the region, particularly in the larger economies of Japan, China and Korea. So, in the imme-

diate future, it is the actions of individuals, such as Cheng and Kim, and their links with venture funds, venture businesses, entrepreneurs and scientists in the West, that are likely to bear the most fruit. □

David Swinbanks is Asia-Pacific editor of Nature.

Taiwan targets biotechnology

Although Taiwan has shown great success in establishing venture businesses in information technology, a push into biotechnology at Hsinchu science park that began in the early 1990s has been slower to develop. However, some new initiatives by the Taiwan government are expected to boost the growth of start-ups in the years ahead.

To date there are only 15 biotech companies in Taiwan's two science parks — Hsinchu near Taipei and Tainan in southern Taiwan — employing a total of fewer than 500 people. About five companies are formed each year, but this figure is expected to double soon, according to Jeff Wang, science adviser at the Ministry of Economic Affairs.

Venture capitalists in Taiwan have been reluctant to invest in biotechnology because of their desire for the quick returns seen in the information technology sector. So, as in Singapore, the Taiwan government has created a large fund of US\$600 million to

promote biotech start-ups over the next five years, and is taking measures to create the necessary infrastructure, including clinical trials, training, and venture funds.

The government is also making substantial investments in research, which are expected to help cultivate start-ups. Some NT\$1,663 million (US\$51 million) are being invested in agrobiotechnology research between 1998 and 2004, while medical genetics will receive NT\$875 million over the same period. Pharmaceuticals and biotechnology research will be given NT\$2,920 million in 1999–2004.

Wang says there are many Taiwanese bioscientists in the West willing to return to Taiwan if and when the environment is right. These programmes, due to employ 430 principal investigators, 630 postdoctoral research fellows and hundreds of PhD students and part-time staff, may mean that now is the time. **D.S.**

End of the brain drain could be in sight

Potter Wickware

Does scientific and engineering talent flow as freely as capital between the world's regions of economic activity and opportunity? Many countries, whether developing ones like Taiwan and Malaysia, or industrialized countries like Germany and Canada, base their economic development policies on the idea that it does, or at least that it could. Hence they channel state resources to start-up companies and offer inducements to scientists and engineers to staff them, particularly their own expatriates who are working in technology ventures in the United States and other developed countries.

These countries look to the high-tech centres of the United States, such as Silicon Valley in northern California and the Route 128 corridor near Boston, as models for the development of their own high-value, knowledge-based industries. The aim is to reduce their dependence on manufacturing, commodity exports, and other more traditional activities.

Indeed, when William Hewlett and David Packard scraped together \$538 in 1938 to start their electronics venture in a garage in Palo Alto, the Santa Clara Valley (as Silicon

Valley was called then) employed mostly farm workers and was famous for its plums and cherries. Last year Hewlett-Packard had nearly 125,000 employees worldwide and reported revenues of \$47 billion.

But is the Silicon Valley model readily transferable to other countries, particularly ones that have thinner traditions of entrepreneurship, or those governed by centralized regimes which have emphasized state planning over individual initiative? Foreign-born scientists and engineers who work in Silicon Valley seem to share the view that an earlier pattern from the nineteenth and early twentieth century still prevails to some extent. That is, move to the United States, become assimilated, and don't look back.

Globe-trotting entrepreneurs

Dinesh Patel exemplifies today's globe-trotting entrepreneurial scientist, who has identified and taken advantage of resources and opportunities as he found them. Born in Zambia, he studied in India, then came to the United States in the 1970s, receiving a PhD in pharmacology at the University of Michigan. In 1985 he co-founded TheraTech, in Salt Lake City, a drug-delivery company which earlier this year



Patel: Silicon Valley lures people back.

was sold to Watson Pharmaceuticals in California for \$350 million. Reflecting on the efforts of countries such as Singapore and Japan to produce technology industries like the one he started, by means of state-supported science parks such as Taiwan's Hsinchu centre (see box above), Patel says he is in general not optimistic about their chances. "If they can, the people who go to them tend to come back to the United States. One of the reasons is that, wherever you are, you need the right connections to make things happen, and someone who begins a career in the United States, in Silicon Valley, say, learns a culture and a management style that may not work in the new locale."

Shuo Vincent Liu echoes the idea that for many the trip to Silicon Valley is a one-way voyage. Liu is a software developer at the bioinformatics company Incyte Pharmaceuticals in Palo Alto, California, who graduated from Beijing University, then studied molec-

ular biology at the University of California at Los Angeles, and biomedical research at the University of Southern California, before going to work in the National Center for



Liu: a return to China would not be easy.

Genome Resources, and private industry. Would he return to China to work in a technology venture? "I'd want to look at it very closely, because it takes lots of resources to mount a start-up company. I'd want to look at which ministry is backing the venture, its budget, how they state their goals, who the people are. Bioinformatics borders on pure research and, since China doesn't have a pharmaceutical industry to speak of, it would be hard for China to support a start-up in my field. Of course the situation is somewhat different in computers and electronics."

Quality-of-life issues are also important factors in the desire of foreign-born workers to stay put. Suni Errabali, chief of Frontier Technology, which does information technology consulting for bioscience companies in Silicon Valley, says: "I might consider moving back to India, my country of origin, but nowhere else. But even this would be a challenge for my children in terms of schooling, and employment and social interaction for my spouse. Most places in the world are not as tolerant to diversity as the San Francisco Bay area."



Huang: escaped from postdoc insecurity.

Scientists who choose to work in start-ups face both benefits and risks. Among the former, says L. Royal Huang, is the opportunity to get away from academic research. Huang got his PhD in biophysics from the Roswell Institute in Buffalo, New York, did postdocs at the University of California, San Francisco, and now works as a scientific programmer at Molecular Applications Group, a bioinformatics software company in Palo Alto. "While I would love to do good science in an academic environment, the postdoc process seems to take forever, and the soft-money financial structure that supports most labs also makes things difficult," he says.

Not that start-ups necessarily offer any more security than a university post. Such companies are too often saddled with inexperienced, vacillating managements and high turnover rates among staff. And, although there is no shortage of research

ideas in bioinformatics and biotechnology, young companies often cannot support them adequately with the journal subscriptions, libraries and computing power that a Squibb or Bell Labs can provide without a second's thought.

But Patrick Nicolas, a software developer at Agorics, an Internet security firm in Sunnyvale, California, says: "I have a very strong preference for working in start-ups." Nicolas came to Silicon Valley from France by way of the Middle East and southern California, and has worked in start-ups since 1987. He observes that "in a start-up one gets the opportunity to wear several hats, in my case software engineering and project management. In large organizations it's not easy for a developer to get direct interaction with the customer and get feedback, and sometimes recognition, for a job well done." He too says he plans to stay in the San Francisco Bay area rather than return to France.

Drawn to Silicon Valley

If anything, Silicon Valley is today more a magnet for foreign-born scientists and engineers than ever before. Anna Lee Saxenian, a professor at the University of California, Berkeley, who studies the sociology of Silicon Valley, reports that 27% of the 4,000 companies started there between 1991 and 1996 are headed by Chinese or Indian executives, compared to 13% in the 1980–85 period. Despite the Asia-related economic slowdown, the job market in Silicon Valley continues to be hot, and employers must compete vigorously to attract qualified talent.

Making it even more difficult for employers in Silicon Valley is the culture of 'musical companies', in which new employees move on to another company just as they become seasoned in the first one. And some of the best qualified employees plan to work for someone else only until they can assemble the resources to start their own companies. Employment prospects in the electronics industry are not as good at the moment as in Internet and drugs firms, but Molly Marr, a spokesperson for the Semiconductor Industry Association, in Menlo Park, California, says: "Silicon Valley is still a boom town and will continue to be a world centre for high-tech start-ups."

The continuing shortage in the United States of skilled technology workers, particularly programmers, which is said to be related to the US educational system, can be satisfied partly by shipping work overseas. Off-shore programming centres have sprung up around the world and, in contrast to Silicon Valley's lofty salaries, their programmers may earn as little as \$1,000 a month. Bangalore, in India, is perhaps the best known, but they also exist in St Petersburg, Beijing, Moscow, Thailand and Jakarta. Bill Lilly, who runs Neva Delta Consulting, in Jackson,

Mississippi, which mediates between code-hungry US technology companies and programmers in Russia, says: "It's a growing business. Small start-ups are looking for client-server architecture, database design, scripting for e-commerce. They're looking for expertise, but cost is also a big factor, so they want to ship work overseas if possible."

In most cases, though, core people must work at the company's site, and in the absence of home-grown talent skilled foreigners must make up the deficit. Last year both companies and foreign job-seekers obtained a boon in the form of an expanded H1B visa programme, which will allow an additional 142,500 skilled foreigners to enter the United States over the next three years. Many will probably apply for permanent residency, but their chances may be dim, depending on their country of origin, among other factors.

Liu says: "One thing foreigners ought to watch out for when they apply for permanent residency is the stability of the company that is sponsoring them. If the company fails, so does their application for residency. And it's not trivial to start the process over again. There's a time limit of six years for visa privileges while the process is under way, and the process itself may take three years or more."

Nicolas adds that, in his opinion, the old visa system instituted by former president Ronald Reagan worked well for years, but has now broken down. When H1B visa holders apply for permanent residency they must submit to the old green card system, where applications may far exceed quotas, especially for countries such as China and India.



Nicolas: predicts the end of the brain drain.

Depending on a great deal of future lobbying and voting in Congress, he forecasts that, starting perhaps five years from now, H1B applicants unsuccessful in their applications for permanent residency will be asked to leave. "They'll be frustrated, but will go back and use what they have learned here to start businesses in their home countries. Their governments will be glad to receive them, and the US can't absorb all of them in any case. It'll be the end of the brain drain."

In this incongruous way, the US start-up culture, which initially drew away their best and brightest, may yet satisfy the aspirations of planners in Singapore, China and other countries to build their own Silicon Valleys. □

Potter Wickware is a science writer and publisher at 101 Reed St, Mill Valley, California 94941, USA. e-mail: wickware@worldnet.att.net or potter@mag.com

How to win venture-capital financing

Diane Gershon

Although record sums of venture capital were invested in early-stage companies in the United States last year, in reality it is very difficult to raise money this way. This is particularly true for life-science and physics-based start-ups, where money is available on only a highly selective basis.

"Venture capital has a very useful but specific role, and unless you fit that set of screening criteria you're not an option," says David Gibson, president of X-ray Optical Systems (see box overleaf). He says venture capitalists are often reluctant to invest smaller amounts of money in very early-stage ventures, because it can take as much time and effort to invest \$100,000 as to invest \$10 million.

What investors knew intuitively was backed up by data presented at this year's National Venture Capital Association meeting in Boston by Cambridge Associates, a group that tracks venture-capital fund returns. The group determined that for the past 13 years venture funds that focused solely on investing in information technology companies gave substantially better rates of return than those with a diversified portfolio of IT and life-science companies.

Although these are sobering statistics, Daniel Egger, managing member of Eno River Capital in North Carolina, still believes this is a good time to invest in the life sciences because the sector is out of favour among most venture investors, so competition is lower. Also, the valuation expectations of early-stage life-science companies are currently very modest, given that they are based on present market conditions. The new fund that Egger and a partner manage focuses solely on the life sciences (although broadly defined) and was set up by the North Carolina Biotechnology Center to back very early-stage companies often deemed too 'early' for most venture funds.

"There is no question that this is now a contrary position to be taking in the venture community," says Egger, who took a slightly unconventional path himself into venture investing, having studied law at college, and formed his own software company, before becoming an 'angel' investor. As a venture fund manager now, he sees part of his job as translating science for investors and translating investor-speak for scientists, and says he can relate to what people go through since he started and later sold his own company.

So how do you convince venture capitalists that you're the only game in town when most investors look at 30, 50 or even 100 business plans for every one that they fund? Perhaps Tracy Lefteroff, worldwide partner-in-charge of Price Waterhouse Cooper's Life

Sciences Industry practice, put it best when he described venture capitalists as "trained sceptics". "While venture capitalists are itching to invest, they are not like 10-year-olds with an allowance burning a hole in their pockets. Venture capitalists have discipline. They can wait for a good deal," he wrote. If you are serious about raising money this way, you can, but you have to commit to it in much the same way that a runner commits to training for a marathon, says Egger. In other words, go for the long haul.

Investment trends

A notable trend in recent years is that the business model of striving to build a fully integrated pharmaceutical company, or FIPCO, is strongly out of favour among life-science venture investors. On the other hand, companies that exhibit the potential to generate products or services, revenues and earnings in the near term are in. Investors no longer believe it is a good use of capital to build all the infrastructure needed to create a FIPCO from scratch, says Egger. And, besides, it has become an almost unattainable goal, given the current state of the capital markets (1998 was marked by a slump in available public capital for life-science companies, particularly in the biotech sector).

Investors want to see licensing and partnering deals with pharmaceutical and biotechnology companies right from the outset, says Egger. Such a business strategy helps to lower the risk profile. Companies that are developing platform or enabling technologies (for example, high-throughput screening systems to generate drug leads for pharmaceutical companies) tend to fare better than those built around a couple of promising molecules.

Writing a business plan

Before you put pen to paper to write a business plan, it is as well to check what resources are available in your area to assist you with the task. In North Carolina, for example, the Council for Entrepreneurial Development offers an evening course called 'fast track' where you can learn just that. Although Egger believes it is important that scientists are involved in writing the plan to familiarize themselves with all facets of the business, it may not be the best use of their time and energy for them to acquire formal business credentials.

What is needed are mentors with business savvy who can shepherd budding entrepreneurs through the early stages of business development, helping them to choose the right business opportunity and bringing people together with complementary skills.

If venture financing is secured, a good early-stage investor should also be fairly 'hands-on', which is largely why venture investors prefer to invest locally.

Become the talk of the town

As a general rule, it is a waste of time and effort to send an unsolicited business plan to a venture investment firm. This is a "relationship business", says Egger, where people go largely on referrals and where it is often the company founders and not the business plan that sell an investor on a business. So get people talking about you and your idea, says Egger, as "even the best [business] plan on paper can't tell you whether the people have the integrity, the intensity, the drive... to stick with such a difficult project and make a go of it".

And if you find the notion of the hard sell distasteful, and many academics still do, then what you need is a buddy, says Egger, who doesn't find it distasteful. Many academics who prefer to remain on the scientific side of the enterprise, or who retain links with academic institutions, choose to team up with a non-academic partner who can drive the process of company development forward full time. You can do it yourself, but it takes people with a lot of energy and drive who are nonetheless realistic about what it takes to succeed. You also need to be something of an extrovert.

One of the personal qualities that all too often is missing, and which sets successful entrepreneurs apart, is what Egger calls "situational awareness". In short, it means understanding your market and where you stand in relation to competing technologies. It also means being able to articulate why your technology is better.

Planning an exit strategy on entry

Perhaps where the cultural gap is widest between venture capitalists and entrepreneurs is in the sense of needing to define an exit strategy from the investment deal almost from the outset, which will give venture capitalists a healthy return on their investment. "It's like getting married and planning the divorce at the same time," says Egger. Academics tend to be locked into a grant-writing mentality of "well, I'll just get another grant next year", but venture capitalists have no choice but to liquidate their venture funds within a set contractual period and to distribute the assets to their investors.

Whereas venture investors won't necessarily expect most of the companies in their portfolio to have a liquidity event (such as an initial public share offering, merger or buy-out) at a handsome profit within three or four years, they will want to see some heart-

felt commitment on behalf of entrepreneurs to do everything possible to achieve such an event. After all, venture-capital investing is a cyclical business of raising money for a new fund, making investments, managing those investments, and then liquidating the fund. So, one excellent question to ask a venture capitalist is, 'where is the venture fund in that cycle?'. A negative response from a venture investor may simply mean that he or she does not have money to invest at that time. □

Diane Gershon is assistant editor, new technology, at Nature Medicine. e-mail: techmed@earthlink.net

Careers and recruitment in Nature

The Careers and Recruitment features published in *Nature* are collected on our website (www.nature.com). Recent features include:

Physics	18 Mar 1999
Canada's expanding knowledge base	11 Feb 1999
Geosciences	3 Dec 1998
Academic-corporate partnerships	24 Sep 1998
Far East emergent economies	6 Aug 1998
Alternative careers for scientists	4 Jun 1998

Cancer research	26 Mar 1998
Cell biology	19 Feb 1998
Careers for women	13 Nov 1997

Future features will include opportunities in chemistry and clinical research, together with a survey of senior appointments and of science in the Benelux countries. Suggestions for topics to be covered within these subjects can be sent to Maxine Clarke at m.clarke@nature.com.

Funding brings high-risk technologies to the marketplace

For start-up companies that find they have few financing options available, funding vehicles such as the Advanced Technology Program (ATP) can prove to be a lifeline. The programme is managed by the US Commerce Department's National Institute of Standards and Technology. "For us, the programme really worked. I honestly doubt if [the company] would be in existence if it hadn't been for the ATP," says Christopher Becker, president and co-founder of GeneTrace Systems in Alameda, California.

Becker and his colleague Joseph Monforte founded GeneTrace Systems in 1994 to develop DNA sequencing systems that combine DNA probing, sequencing and sizing reactions with laser-based time-of-flight microscopy. The aim is to develop analysis methods that are faster and cheaper than gel-based approaches. Since the company received ATP funds of about \$2 million, GeneTrace has entered into equity and licensing agreements with Incyte Pharmaceuticals and Monsanto, and now has 60 staff.

Becker's sentiment seems to be echoed by a report¹ that evaluated the first group of completed ATP-funded research projects. The report provides an assessment of the 38 ATP projects completed by March 1997 (12 projects were terminated before completion). Two-thirds of the companies would not have proceeded with their projects without ATP funding, while the rest reported gains of 18 months or more in product development as a result of the ATP award (see Table 1).

Typically, ATP funds high-risk, pre-competitive, generic technologies that have

the potential to produce significant commercial payoffs and benefits to the US economy. Although any size of company can apply, more than half the awards go to individual small businesses or to joint ventures led by a small business. Awards are made after a rigorous competitive review process that assesses the technical merit and economic potential of the project. ATP does not provide funds for basic research or for product development, and has funded an average of 12% of industry proposals received since 1990 when the programme began. Funding for single-applicant companies and joint ventures is on a cost-sharing basis. Single-applicant companies are expected to cover all the indirect project costs, although ATP may cover up to 100% of direct costs.

"The step between 'is the theory real?' to 'can you do it practically?' is where ATP typically sees its role," says David Gibson of X-ray Optical Systems in Albany, New York. The company designs optical systems that bend and focus X-rays, which have potential applications in fields as diverse as medical imaging, material analysis, X-ray lithography and astronomy. The company received almost \$2 million from ATP, which took it to the point where it could make X-ray optics reliably and repeatedly to specification. Although it was not yet commercial, it had begun to explore ways to use the optics in specific applications. "Most of the technical risk was gone," says Gibson.

X-ray Optical Systems' financing story has an all too familiar ring about it among start-ups. The company was founded by two scientists, one from the United States and the

other from the Soviet Union. Gibson, a former management consultant with an engineering and economics background, was brought in initially to do a six-month feasibility study to determine whether the technology would work, and whether the company could create economic value with it and then capture that value. Before the company secured ATP funding in the early 1990s, it was financed largely by its founders, who mortgaged their homes, borrowed on credit and then from friends and family.

Gibson says venture capital was not an option at the time because of the degree of technical risk involved and because investors tend to shy away from physics-based technologies. Also, although the company had the potential to create a lot of value, the venture capitalists correctly identified that it was not going to be able to capture all that value because it would be in the business of making components for sale to end users and equipment manufacturers.

It was for these same reasons that the company's project was a good fit for ATP, although its application was turned down first time around. Gibson says they did look into the possibility of raising money through strategic partnering but, because the X-ray optics had such a broad range of potential applications, it was hard to identify a potential suitor. Although the company probably could have raised money that way, Gibson believes that at the time the founders would probably have had to give up a majority stake in the company to do so, and it would most likely have been a non-US company that would have been interested.

"ATP allowed us to keep it in the US and to keep it independent," he says. And ATP funding provided technical validation that has unlocked some private money, says Gibson. Although the company is now at a point when perhaps it could consider venture capital, "it's still not clear that it would be a good choice for them or us," he says.

For more information about ATP, see <http://www.atp.nist.gov> **D.G.**

1. Long, W. *Advanced Technology Program: Performance of Completed Projects — Status Report No. 1*. Special Publication 950-1 (NIST, Gaithersburg, Maryland, 1999).

Table 1 Impact of ATP funding on conducting projects

Would have proceeded without ATP funding	Number of ATP projects	%
Yes, but at a slower pace	11	34
with a delay of:		
18 months	4	
21 months	3	
24 months	3	
60 months	1	
No	21	64